Machine Learning for Diagnosing Drug Users and Types of Drugs used

Anthony Anggrawan[1]
Lalu Ganda Rady Putra[5]
Computer Science Study Program
Bumigora University
Mataram, West Nusa Tenggara, Indonesia

Christofer Satria[2]
Visual Communication Design Study Program
Bumigora University
Mataram, West Nusa Tenggara, Indonesia

Che Ku Nuraini[3]
Department of Centre of Research and Innovation Management
Universiti Teknikal Malaysia Melaka, Malaysia

Lusiana[4]
Technical Information Study Program, STMIK AMIK Riau
Riau, Indonesia

***Abstract*— Drug use is very detrimental to the physical and psychological health of users. Drug abuse also causes addiction and is a global epidemic. Therefore it is not surprising that scientific research related to drugs has attracted attention for research. However, many factors become obstacles in the medical services of the drug user, including cost, flexibility, and a slow process. Meanwhile, electronic systems can speed up handling time, improve work efficiency, save costs and reduce inspection errors. It means that a breakthrough is needed in developing a platform that can identify drug users. Therefore, this research aims to build machine learning with expertise like an expert who can diagnose drug users and distinguish the types of drugs used by drug users. The expert system on machine learning was developed using the Forward Chaining and Certainty Factor methods. This study concludes that the expert system on machine learning developed can be used to diagnose drug users and distinguish the types of drugs used with an accuracy of up to 80%. The implications of the expert system on machine learning are an alternative method for narcotics officers and medical doctors in diagnosing drug users and the types of drugs used.**

Keywords—Machine learning; Drug; Expert system; Forward chaining; Certainty factor

I. INTRODUCTION

Social environment factors have influenced others to engage in drug use [1] [2]. Adult figures who are addicted to drugs have a great influence on the behavior of others to become addicted [3]. Poor, low skills, life pressure, anxiety, and deviant behavior are factors that also lead to drug use [3]. Relaxing, drinking, staying up late, increasing enthusiasm, and relieving stress are other triggering factors for many drug use among young people [4]. Drug use, including amphetamines, marijuana, cocaine, heroin, and the like, are a major public health problem in physiological symptoms, resulting in behavioral changes, cognitive problems, and mental health [5]. Drug use also affects the physiology and behavior of future generations [2].

Drug abuse causes physical dependence (addiction) or relapse to continuously consume [6], although it has resulted in physical and psychological problems [6]. The previous research show that drug users are very high [7] [1] and increasing globally [8]. Drug abuse has become a global epidemic that affects human behavior [9]. Because drug use is very detrimental to the physical and psychological harm of the user, it is not surprising that this research related to drugs has attracted attention for scientific research [10].

Another factor that is often considered in medical services is the cost factor and its inflexibility (Bevan & Patel, 2016). Processes that are done manually tend to cause delays in medical diagnosis [11]. Using an electronic system can speed up handling time, improve work efficiency, and save costs [11]. Using an electronic system allows lower errors and eliminates omissions in deciding the test results and achieving the results [12]. The success of curing drug abuse and dependence is still limited; this includes the lack of success in the early identification of at-risk populations, resulting in increased death rates due to overdose [13]. In other words, not paying attention to the early symptoms of consuming drugs will be disastrous and make people who are loved suffer the destructive effects of the substance [14].

Meanwhile, if it turns out to be able to identify it early, it can prevent harmful consequences in the future that are sure to occur [14]. It means that there is a need for breakthroughs in developing platforms that can identify and screen patients susceptible to addiction after using opioid drugs [13]. Therefore, this research aims to develop a machine learning that has expertise like an expert. The expert system created can identify and screen or diagnose early drug users and the types of drugs used by using the Certainty Factor and Forward Chaining methods. The certainty factor method measures the certainty of the type of drug used by the user or patient who conducts consultations. On the other hand, forward chaining plays a role in the flow of the reasoning process from beginning to end based on data mining of physical symptoms of drug users and the types of drugs used (collected or explored previously).

Medical data is helpful as the knowledge that helps make scientific decisions regarding drug use [15]. Electronic medical use based on doctor's notes is useful for an effective treatment medium [10]. In the meantime, data mining is capable of electronic checks based on the patient's medical record [10].

Besides, the machine learning methods are a technique that can be useful for finding correlations based on the case for prediction purposes [16]. Unfortunately, machine learning is still few in the medical field due to technical problems [17]. Therefore, the simple machine learning method built in this research by imitating (studying) human knowledge in analyzing the physical symptoms of drug users and then implementing it in predicting drug users and identifying drug types used by users. It means that the expert system in machine learning has intelligence like an expert in diagnosing users and the types of drugs used from the physical symptoms that arise from drug users. Taking into account that the current use of information and communication technology (ICT) is growing or expanding very quickly or booming [18]. Therefore, the embodiment of the machine learning system in this research is website-based. So, anyone (the public) can access it from anywhere and has flexibility because it can work on various devices and operating systems. A machine learning system in this research is helpful for early diagnosis without having to examine a narcotics laboratory and without a doctor or expert.

It is necessary to know the percentage of machine learning efficacy in identifying drug users and the type of drug used. It means that further testing to determine the actual percentage of machine learning efficacy still needs to be done. This study makes this happen by comparing the test results achieved by machine learning based on symptoms of drug users compared to the test results achieved from laboratory tests of drug users' urine in identifying drug users and the type of drug used. If machine learning has high efficacy, it can save time and cost of drug testing for suspects or drug users by using machine learning compared to testing drugs on urine or blood for suspects and drug users.

Some recent works related to this research:

- Zhongheng Zhang (2016) introduced the k-nearest neighbor (kNN) method as a simple machine learning method for modeling [17]. The similarity between the research in this article and the previous one is that they both use a simple approach to machine learning. While the difference is that the research uses the certainty factor method and forward chaining for machine learning, while previous research uses the kNN method for machine learning. Another difference is that the previous research was focus on predicting the class from the new dataset to the most similar class. In contrast, the research in this article focuses on machine learning to diagnose drug users and the types of drugs used.
- Anthony Anggrawan, Khasnur Hidjah, and Jihadil Qudsi S. (2017) implement intelligent application programs to detect kidney failure [19]. The previous research and the research in this article have similarities in developing web application programs with the PHP programming language and MySQL database. In addition, the last analysis used medical data on failure cases to diagnose new renal illness issues using CBR (Case-Based Reasoning method). In contrast, the articles in this study use the expertise of drug experts (specialists) as knowledge of the application system for early diagnosis of drug users and the types of drugs used by drug users using the Forward Chaining and Certainty Factor methods.
- Kurnia Muludi, Radix Suharjo, Admi Syarif, and Fitria Ramadhani (2018) identified tomato plant diseases [20]. This previous research and the research in this article both implements forward chaining and certainty factor methods. However, the difference in the last research is to build an expert system to identify plant diseases based on android [20], while the research in the article builds an expert system to identify users of drugs and the types of drugs used based on the website.
- Munaiseche, Kaparang, and Rompas (2018) built an expert system to assist doctors in diagnosing eye diseases [21]. In contrast to the research in this article, it is to build an expert system to diagnose drug users and the types of drugs used. Furthermore, this previous research used the forward chaining method, while the research in this article uses the forward chaining method and certainty factor. The similarities between the previous study and the research in this article are that both use PHP and MySQL in building an expert system.
- Ninive Von Greiff and Lisa Skogens (2021) investigates a drug user recovery program for drug addiction [22]. The research method is the interview or qualitative approach [22]. The similarity of the research in this article with previous studies is that they both study drug users. The difference is in previous studies examining the results of addiction recovery on drugs with the interview method. Meanwhile, the research in this article builds machine learning that has an intelligent system to detect drug users and the types of drugs used.

The latest related work identifies that the article in this study has a novelty that no previous researcher has researched. Other strength of this research is conducting a comparative test to determine the efficacy of machine learning or expert systems developed in identifying users and the types of drugs used by users that have not been studied before.

The systematics of writing this paper is as follows: the following sub-section discusses the research methodology, which includes research data and research methods used. The next subsection discusses the results and discussion of the research. Finally, the conclusions obtained from the study results and suggestions for further research are narrated in the Conclusions subsection.

II. Research Methodology

This study is a case study at the Indonesian National Narcotics Agency (*Badan Narkotika Nasional* or BNN) in Mataram, Indonesia. The number of drug users used as samples to test the expertise and accuracy of the machine learning built in this study was 30. The selection of data samples in this study was random. This research's development of machine learning expertise consists of stages: knowledge acquisition, expert system design (programming), machine learning/expert system testing, and accuracy test (see Fig.1).

A. Knowledge Acquisition

For the system development stage, the effort made is to obtain knowledge from drug experts, which is used as a knowledge base to build the machine learning or expert systems. The method used in obtaining knowledge related to narcotics is the interview method. The knowledge gained is the

knowledge about the types of drugs and their symptoms. Based on the knowledge obtained, there are ten types of drugs that drug users dominantly use, and there are 27 types of drug symptoms

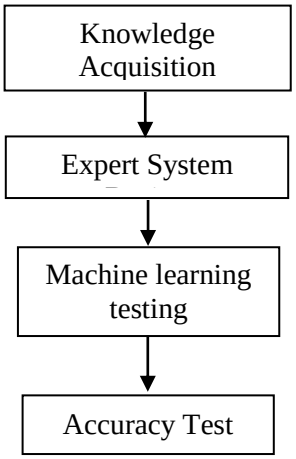

Fig.1. Stages in the development of an expert system on machine learning

In the knowledge acquisition stage, an expert from the Indonesian National Narcotics Agency (in Mataram, Indonesia) shares knowledge about drug use, including those related to symptoms and the types of drugs used by drug users. The knowledge gained at this stage serves as a knowledge base in building expertise from machine learning.

B. *Expert System Design*

The expert system design stage is a process for modeling the data that has been collected and designing an application system that is planned according to programming problems and the acquisition of knowledge obtained. This stage in computer science is known as planning the use case diagram design, data flow diagram (DFD) design, database design, and flowchart to be built on the application program. The programming stage is the implementation stage of the system design plan into a computer programming language. This research uses the PHP and MySQL programming languages. The computer application program that is built is a cloud application program. Application development with PHP programming language and MYSQL database makes application programs can run via the web. By being stored on the server computer, the application program is ubiquitous. The ubiquitous application program means that the application program can be accessed from anywhere and at any time [23].

C. *Machine learning testing*

The machine learning testing phase is the functional testing phase of the built application or black-box testing. Black-box testing is a test that no longer involves programming code or programming languages. In short, black-box testing on the expert system in this study is to determine whether the expert system built is under the list of desired system requirements.

D. *Accuracy Test*

Testing accuracy in machine learning is to determine the level of expertise of the expert system built in this study

III. Result and Discussion

A. *Knowledge Acquisition*

The knowledge acquisition stage is the stage of acquiring the required knowledge data. The acquired knowledge acquisition data is useful in solving programming logic in diagnosing users and the types of drugs used by drug users. Table I shows the code for the type of drug used by drug users. Meanwhile, Table II presents the code of symptoms caused by drug users.

TABLE I. List of types of drugs

No.	Drug Type Code	Drug Name
1	P001	Cocaine
2	P002	Marijuana
3	P003	Ecstasy
4	P004	Heroin/Putaw
5	P005	Methamphetamine
6	P006	Hallucinogen
7	P007	Amphetamine
8	P008	Pethidine
9	P009	Codeine
10	P010	Morphine

TABLE II. List of Symptom of Drug User

Symptoms of Drug User			
Code	**Symptom**	**Code**	**Symptom**
G001	Out of breath	G015	Difficult to focus
G002	Anxious and restless	G016	Difficult to rest
G003	Nausea and vomiting	G017	Weight loss
G004	Diarrhea	G018	Dry mouth
G005	Convulsions	G019	Blurred vision
G006	Easy to get angry	G020	Changes in skin color
G007	Depression	G021	Constipation
G008	Changes in sleep patterns	G022	Stomachache
G009	Sweating	G023	Drowsiness
G010	Chills (Hot cold)	G024	Itching
G011	Shaking	G025	Difficulty urinating
G012	Insomnia	G026	Mood swings
G013	Fast heart rate	G027	Dizziness
G014	Increased blood pressure		

TABLE III. Rule Base of Types and Symptoms of Basic Diagnosis

Rule Base	
Drug Type	Symptom
P001	G001 and G002 and G003 and G004 and G005
P002	G006 and G002 and G007 and G008 and G009 and G010
P003	G005 and G011 and G012 and G013 and G014
P004	G015 and G002 and G007 and G016
P005	G011 and G001 and G016 and G017 and G018
P006	G009 and G011 and G018 and G019 and G010 and G014
P007	G018 and G003 and G004 and G005 and G001 and G020
P008	G007 and G013 and G005 and G003
P009	G027 and G003 and G018 and G021 and G022
P010	G023 and G024 and G009 and G025 and G026

After modeling the acquired knowledge acquisition data or knowledge representation (as shown in Table III) the next step is to implement it into the certainty factor algorithm. The certainty factor uses a value between 0.2 and 1.0 to assume a level of confidence in the data. A simulation of the calculation of the certainty factor was carried out based on the weight of symptoms arising from the type of drug used by drug users with weights of 0.8 and 1.0 according to the opinion of the drug expert (See Table IV).

TABLE IV. DETERMINATION OF DRUG SYMPTOM WEIGHT SCORE ACCORDING TO THE DECISION OF DRUG EXPERTS

No	Symptom	Weight Score
1	Very often	1
2	Often	0,8
3	Never	0

So, on the drug symptom weighted score given to the certainty factor, a score of 0 indicates that drug users do not experience these symptoms. If a drug user experiences symptoms, then the weighted score given for the frequently experienced symptoms is 0.8 and the most frequently experienced is 1.0, according to the drug expert's decision.

TABLE V. KNOWLEDGE BASE

Symptom	J-001	J-002	J-003	J-004	J-005	J-006	J-007	J-008	J-009	J-010
	CF	CF	CF	CF	CF	CF	CF	CF	CF	CF
Out of breath					0,8		1			
Anxious and restless	0,8	0,8		1						
Nausea and vomiting	0,8						0,8	0,8	0,8	
Diarrhea	1						0,8			
Convulsions	0,8		0,8				0,8	0,8		
Easy to get angry		1								
Depression		0,8		0,8				1		
Changes in sleep patterns		0,8		0,8						
Sweating		0,8				0,8				0,8
Chills		0,8				0,8				
Shaking			0,8		0,8	0,8				
Insomnia			0,8							
Fast heart rate			0,8					0,8		
Increased blood pressure			0,8			1				
Difficult to focus				1						
Difficult to rest					0,8					
Weight loss					0,8					
Dry mouth					1	0,8	0,8		1	
Blurred vision						0,8				
Changes in skin color							0,8			
Constipation									0,8	
Stomachache									0,8	
Drowsiness										1
Itching										0,8
Difficulty urinating										0,8
Mood swings										1
Dizziness									0,8	

This study's knowledge base of machine learning expertise is the symptoms, types of drugs, and CF rules obtained from drug experts (see Table V). The knowledge base is an essential component that contains the knowledge possessed by competent experts in the related field (i.e., narcotics in this study). Furthermore, the knowledge base is the basis for decision-making in an expert system, where this decision-making is related to the process of retrieving previously collected and stored knowledge.

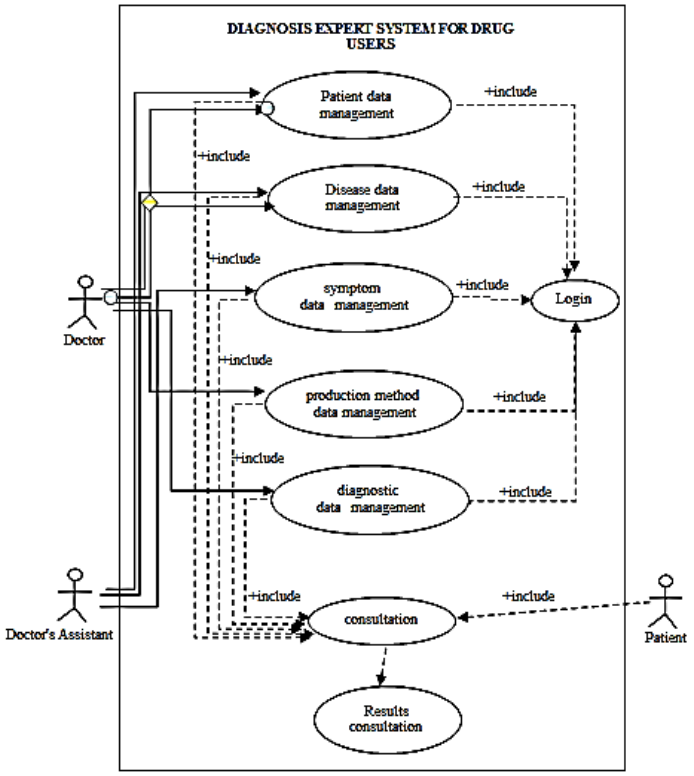

Fig. 2. Use Case Diagram on machine learning

This study has a database that stores records of users, patients, symptoms, and types of drugs, including diagnostic data, so it is necessary to design a data workflow model to realize a structured program.

The Data Flow Diagram (DFD) in Fig. 3 and Fig. 4 illustrate where the data flow comes from and where the data processing on the expert system is built. The context diagram in Figure 3 shows the data flow of the system globally. In contrast, the overview diagram in Figure 4 shows a more detailed data flow that the system performs and its engagement with external data.

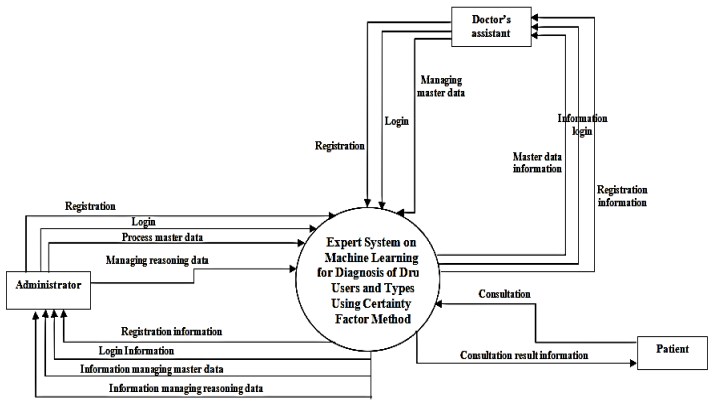

Fig. 3. Context Diagram of Data Flow on Machine Learning

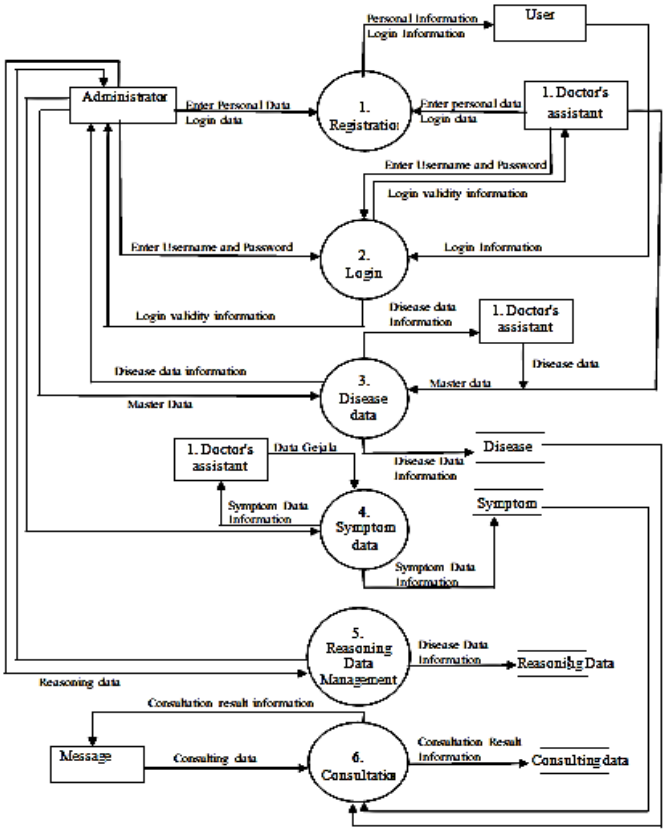

Fig. 4. Overview Diagram of Data Flow on Machine Learning

The flow diagram in fig. 5 shows a series of flow relationships in the expert system built in this study or shows the overall process sequence in building an expert system in this study.

The flow diagram contains a more detailed description of how each step of the procedure is actually carried out in building an expert system on machine learning that can diagnose users and the types of drugs used by users.

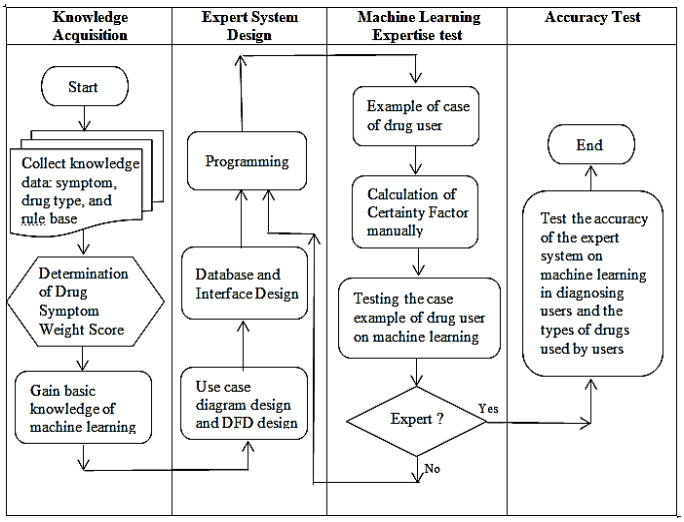

Fig. 5. Flow diagram of the whole Process of Building an Expert System on the Machine Learning

B. Machine learning testing

Expertise testing of machine learning is carried out using case samples from a patient. For example, in one case, a drug patient had symptoms of shortness of breath, depression, chills, anxiety and restlessness, and irritability. Symptoms of drug patients who have symptoms of shortness of breath, depression, chills, anxiety and restlessness, and irritability are symptoms of drug users: (1) Cocaine, (2) Cannabis, (3) Heroin, and (4) Amphetamine.

The formula for CF is:

CF [H,E] = CF [H] * CF [E]

CF Combine CF [H,E]1 = CF [H,E] 1 + CF [H,E] 2 * (1-CF [H,E] 1)

CF Combine CF [H,E] old3 = CF [H,E] old + CF[H,E] * (1-CF[H,E] old)

Based on manual calculations, the results are as follows:

1. For J-001 = Cocaine

   G001 = Out of breath (0.8)

   CF [H,E] = CF [H] * CF [E]

   = 0.8 * 0.8

   = 0.64

   G002 = Anxious and restless (0.8)

   CF [H,E] = CF [H] * CF [E]

   = 0.8 * 0.8

   = 0.64

   CFk1 = CF [H,E]1 + CF [H,E]2 * (1-CF[H,E]1)

   = 0.64 + 0.64 * (1-0.64)

   = 0.870

   So the expert CF from the symptoms entered by the user for the type of drug Cocaine is probably 0.870 or 87%.

2. For J-002 = Marijuana

   G007 = Depression (0,8)

   CF [H,E] = CF[H] * CF[E]

   = 0.8 * 0.8

   = 0.64

   G010 = Chills (0.8)

   CF [H,E] = CF[H] * CF[E]

   = 0.8 * 0.8

   = 0.64

   G002 = Anxious and restless (0.8)

   CF [H,E] = CF[H] * CF[E]

   = 0.8 * 0.8

   = 0.64

   G006 = Easy to get angry (1)

   CF [H,E] = CF[H] * CF[E]

   = 0 * 1

   = 0

   CFk1 = CF [H,E]1 + CF [H,E]2 * (1-CF[H,E]1)

   = 0.64 + 0.64 * (1-0.64)

   = 0.870

   CFk2 = CFk1 + CF[H,E]3 * (1-CFk1)

= 0.870 + 0.64 * (1-0.870)

= 0.953

CFk3 = CFk2 + CF[H,E]4 * (1-CFk2)

= 0.953 + 0 * (1- 0.953)

= 0.953

So the CF of the symptoms entered by the user for the type of marijuana drug is likely to be 0.953 or 95%.

3. For J-004 = Heroin

G007 = Depression (0.8)

CF [H,E] = CF[H] * CF[E]

= 0.8 * 0.8

= 0.64

G002 = Anxious and restless (1)

CF [H,E] = CF[H] * CF[E]

= 0 * 1

= 0

CFk1 = CF[H,E]1 + CF[H,E]2 * (1-CF[H,E]1)

= 0.64 + 0 * (1-0.64)

= 0.64

So the CF of the symptoms entered by the user for the type of heroin drug is most likely 0.64 or 64%.

4. For J-007 = Amphetamine

G001 = Out of breath (1)

CF[H,E]1 = CF[H]1 * CF[E]2

= 0.8 * 1

= 0.8

So the CF of the symptoms entered by the user for the type of Amphetamine is most likely 0.8 or 80%.

Based on the value of manual calculations, the largest CF value is taken, which is 0.953 or 95% with the type of marijuana drug. It means the patient is using a type of marijuana drug. A case example is tested on an expert system application program (or on machine learning). If the patient's symptoms in the case sample (with the same symptoms) are entered into the expert system built in this study, the result of the process is as shown in Fig. 6.

HASIL KONSULTASI

Nama	:	Andi
Umur	:	24
Jenis Kelamin	:	Laki-laki
Pekerjaan	:	PNS
Alamat	:	ampenan

No	Pertanyaan	Jawaban
1	Sesak Nafas	TIDAK
2	Cemas dan Gelisah	IYA
3	Mual dan Muntah	TIDAK
4	Diare	TIDAK
5	Kejang-kejang	TIDAK
6	Mudah Marah	TIDAK
7	Depresi	IYA

Fig. 6. Screenshot of expert system questions about drug symptoms experienced by patients

Expert system testing on machine learning shows that the expert system has succeeded in correctly identifying the user and the type of drug used by the user. In order to know how accurate the machine learning expertise is, this study also tested several other patients by comparing the results with the urine test results at the Indonesian National Narcotic Agency laboratory in Mataram, Indonesia.

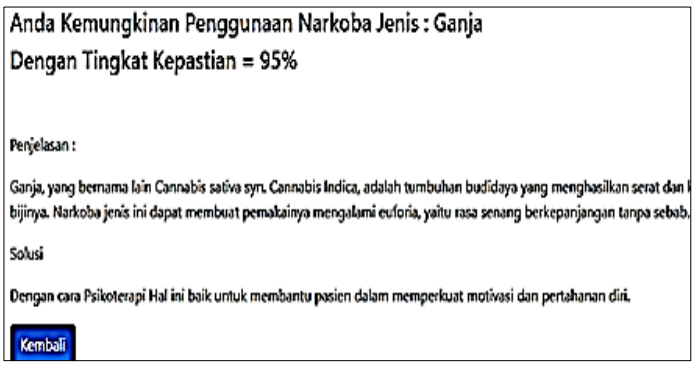

Fig.7. Screenshot of the expert system test results on the type of drug used by the patient

Fig. 7 describes it as follows: You are probably using a type of drug with a 95% certainty. Another narrative in figure 7 is: another name for cannabis is cannabis sativa. Cannabis sativa is a cultivated plant that contains fiber and narcotic substances in its seeds. This drug makes user experience euphoria, namely a prolonged feeling of pleasure for no reason. The cure is psychotherapy, which helps the patient strengthen the motivation to stop using it.

Expert system testing on machine learning shows that the expert system has succeeded in correctly identifying the user and the type of drug used by the user. In order to know how accurate the machine learning expertise is, this study also tested several other patients by comparing the results with the urine test results at the Indonesian National Narcotic Agency laboratory.

C. Accuracy Test of Machine Learning

The accuracy test of machine learning in this study is to determine the expert performance of the application system built-in diagnosing users and the types of drugs used. Testing the level of accuracy of machine learning expertise is to compare the suitability of the results with the urine test results from patients at the Indonesian National Narcotics Agency. In a trial of 30 times on 30 patients, there were 30 results of machine learning tests that can correctly identify drug users and as many as 24 machine learning test results that can detect the types of drugs used by drug users. It means that the test results on the data of 30 drug patients show that an expert system on machine learning built using the Certainty Factor method has expertise in diagnosing drug users up to 80 percent. The accuracy rate of up to 80 percent is obtained from the results of the calculation of 24 divided by 30 and multiplied by 100%. Table VI shows the details of the test results in diagnosing the type of abuse of drug users.

TABLE VI. Machine Learning Expertise Accuracy Test Result

No	Case	System Result	Expert Result	Suitability
1	G001,G002,G003,G004,G005	Cocaine	Cocaine	Suitable
2	G006,G002,G007,G008,G009,G010	Marijuana	Marijuana	Suitable
3	G005,G011,G012,G013,G014	Ecstasy	Ecstasy	Suitable
4	G015,G002,G007,G016	Heroin/Putaw	Heroin/Putaw	Suitable
5	G011,G001,G016,G017,G018	Methamphetamine	Methamphetamine	Suitable
6	G009,G011,G018,G019,G010,G014	Hallucinogen	Hallucinogen	Suitable
7	G018,G003,G004,G005,G001,G020	Amphetamines	Amphetamines	Suitable
8	G007,G013,G005,G003	Pethidine	Pethidine	Suitable
9	G027,g003,G018,G021,G022	Codeine	Codeine	Suitable
10	G023,G024,G009,G025,G026	Morphine	Morphine	Suitable
11	G003,G006,G007,G0015	Pethidine	Pethidine	Suitable
12	G001,G008,G009,G0018	Codeine	Codeine	Suitable
13	G001,G002,G005,G009,G011,G015,G018	Codeine	Hallucinogen	Not suitable
14	G004,G0012,G013,G017	Ecstasy	Ecstasy	Suitable
15	G008,G011,G017,G018,G019	Methamphetamine	Methamphetamine	Suitable
16	G001,G013,G015,G020,G022	Amphetamines	Amphetamines	Suitable
17	G017,G020,G021,G022,G023G025,G027,	Codeine	Morphine	Not suitable
18	G002,G007,G011,G015,G016	Pethidine	Heroin/Putaw	Not suitable
19	G003,G007,G010,G014,G019	Pethidine	Pethidine	Suitable
20	G008,G011,G015,G018,G019,G020G022,G023	Codeine	Amphetamines	Not suitable
21	G002,G006,G007,G008,G009,G010,G019,G027	Marijuana	Marijuana	Suitable
22	G005,G011,G012,G014,G019,G020	Hallucinogen	Ecstasy	Not suitable
23	G001,G006,G007,G0010,G024	Pethidine	Marijuana	Not suitable
24	G003,G007,G009,G021	Morphine	Morphine	Suitable
25	G005,G012,G017,G022	Ecstasy	Ecstasy	Suitable
26	G007,G012,G015,G026	Heroin/Putaw	Heroin/Putaw	Suitable
27	G002,G008,G015,G023	Heroin/Putaw	Heroin/Putaw	Suitable
28	G006,G009,G015,G026	Morphine	Morphine	Suitable
29	G007,G011,G018,G023,G027	Codeine	Codeine	Suitable
30	G008,G015,G025,G026	Morphine	Morphine	Suitable

IV. CONCLUSION

The results of this study found that: (a). Machine learning in this study can predict drug users and types of drugs based on the symptoms that drug users complain about. (b). This study machine learning acquired knowledge about the symptoms of drug users, types of drugs, and basic knowledge related to the weight of the certainty factor of each type of drug and the symptoms caused so that it can diagnose drug users and the types of drugs used by users. (c). The accuracy of machine learning achieved in this study in predicting the types of drugs used by users and the types of drugs used by users reached 80%. (d). The expert system in this research is website-based, so that the expert system from this research can be used by various parties and in different places to identify users and the types of drugs used by users.

The implication of this research result is that the expert system built in this study can be a tool (choice) to replace or complete the testing system for drug users through urine testing in the laboratory.

The drawback of the results of this study is that machine learning expertise in this study is only limited to simple machine learning, as is the case with simple machine learning which was built on previous research by Zhongheng Zhang (2016), which used the KNN method in building learning machines. Furthermore, the machine learning expertise generated from this research is only limited to the expertise possessed in accordance with the knowledge obtained (symptoms, types of drug abuse, rule base, and calculation of certainty factor) under study. Therefore, further research needs to build machine learning that can increase its expertise based on more new data and use another method.

REFERENCES

[1] R. Jiménez, J. Anupol, B. Cajal, and E. Gervilla, "Data mining techniques for drug use research," *Addict. Behav. Reports*, vol. 8, pp. 128–135, 2018.

[2] F. M. Vassoler, E. M. Byrnes, and R. C. Pierce, "The impact of exposure to addictive drugs on future generations: Physiological and behavioral effects," *Neuropharmacology*, vol. 76, no. PART B, pp. 269–275, 2014.

[3] P. K. Shanmugam, "The Influence of Social Factors in Drug Addiction–A Mini Review of Work by Miller & Carroll (2006)," *J. Alcohol. Drug Depend.*, vol. 05, no. 04, pp. 4–6, 2017.

[4] A. Boys, J. Marsden, and J. Strang, "Understanding reasons for drug use amongst young people: A functional perspective," *Health Educ. Res.*, vol. 16, no. 4, pp. 457–469, 2001.

[5] G. López, L. M. Orchowski, M. K. Reddy, J. Nargiso, and J. E. Johnson, "A review of research-supported group treatments for drug use disorders," *BMC Public Health*, vol. 16, no. 51, pp. 1–21, 2021.

[6] Z. Justinova, L. V. Panlilio, and S. R. Goldberg, "Drug Addiction," *Natl. Libr. Medicine*, vol. 1, no. 1, pp. 310–335, 2009.

[7] T. Saah, "The evolutionary origins and significance of drug addiction," *Harm Reduct. J.*, vol. 2, pp. 1–7, 2005.

[8] G. Leshner, E. M.Stevens, S. Kim, N. Kim, T. L.Wagener, and A. C. Villantie, "Cognitive and affective responses to marijuana prevention and educational messaging," *Drug Alcohol Depend.*, vol. 225, no. August, pp. 1–3, 2021.

[9] M. Zaman *et al.*, "Drug abuse among the students," *Pakistan J. Pharm. Res.*, vol. 1, no. 1, p. 41, 2015.

[10] L. W. Chou, K. M. Chang, and I. Puspitasari, "Drug Abuse Research

Trend Investigation with Text Mining," *Comput. Math. Methods Med.*, vol. 2020, pp. 1–8, 2020.
[11] A. Bevan and N. Patel, "An Electronic Prescription Alerting System-Improving the Discharge Medicines Process," *Arch. Dis. Child.*, vol. 101, no. 9, p. e2.55-e2, 2016.
[12] A. Tsyben, N. Gooding, and W. Kelsall, "Assessing the Impact of a Newly Introduced Electronic Prescribing System Across a Paediatric Department – Lessons Learned," *Arch. Dis. Child.*, vol. 101, no. 9, p. e2.13-e2, 2016.
[13] M. Mahmoudi, S. Pakpour, and G. Perry, "Drug-Abuse Nanotechnology: Opportunities and Challenges," *ACS Chem. Neurosci.*, vol. 9, no. 10, pp. 2288–2298, 2018.
[14] J. Redman, "Recognizing the Warning Signs of Drug Addiction : What You Need to Know," *Mental Health and Counseling Studies*. pp. 1–7, 2021.
[15] N. Jojen, "A Survey Paper on Data Mining Techniques in Drug Industry," *Int. J. Eng. Res. Technol.*, vol. 3, no. 30, pp. 296–299, 2015.
[16] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or organ category," *Front. Chem.*, vol. 6, no. May, pp. 1–11, 2018.
[17] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.
[18] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 4, pp. 319–326, 2021.
[19] A. Anggrawan, K. Hidjah, and Q. S. Jihadil, "Kidney failure diagnosis based on case-based reasoning (CBR) method and statistical analysis," in *2016 International Conference on Informatics and Computing, ICIC 2016*, 2017, pp. 298–303.
[20] K. Muludi, R. Suharjo, A. Syarif, and F. Ramadhani, "Implementation of forward chaining and certainty factor method on android-based expert system of tomato diseases identification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 9, pp. 451–456, 2018.
[21] C. P. C. Munaiseche, D. R. Kaparang, and P. T. D. Rompas, "An Expert System for Diagnosing Eye Diseases using Forward Chaining Method," in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 306, no. 1, pp. 1–8.
[22] N. Von Greiff and L. Skogens, "Recovery and identity: a five-year follow-up of persons treated in 12-step-related programs," *Drugs Educ. Prev. Policy*, pp. 1–10, 2021.
[23] A. Anggrawan, N. Ibrahim, S. Muslim, and C. Satria, "Interaction between Learning Style and Gender in Mixed Learning with 40 % Face-to-face Learning and 60 % Online Learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, pp. 407–413, 2019.

Similarity Check _ Machine Learning for Diagnosing Drug Users and Types of Drugs used

By Anthony Anggrawan

WORD COUNT	5295	TIME SUBMITTED	17-OCT-2021 07:47PM
		PAPER ID	77793429

Machine Learning for Diagnosing Drug Users and Types of Drugs used

Anthony Anggrawan[1]
Ni Gusti Ayu Dasriani[5]
Mayadi[6]
Computer Science Study Program
Bumigora University, Mataram, Indonesia

Christofer Satria[2]
Visual Communication Design Study Program
Bumigora University, Mataram, Indonesia

Che Ku Nuraini[3]
Department of Centre of Research and Innovation Management
Universiti Teknikal Malaysia Melaka, Malaysia

Lusiana[4]
Technical Information Study Program
STMIK AMIK Riau, Riau, Indonesia

***Abstract*— Drug use is very detrimental to the physical and psychological health of users. Drug abuse also causes addiction and is a global epidemic. Therefore it is not surprising that scientific research related to drugs has attracted attention for research. However, many factors become obstacles in the medical services of the drug user, including cost, flexibility, and a slow process. Meanwhile, electronic systems can speed up handling time, improve work efficiency, save costs and reduce inspection errors. It means that a breakthrough is needed in developing a platform that can identify drug users. Therefore, this research aims to build machine learning with expertise like an expert who can diagnose drug users and distinguish the types of drugs used by drug users. The expert system on machine learning was developed using the Forward Chaining and Certainty Factor methods. This study concludes that the expert system on machine learning developed can be used to diagnose drug users and distinguish the types of drugs used with an accuracy of up to 80%. The implications of the expert system on machine learning are an alternative method for narcotics officers and medical doctors in diagnosing drug users and the types of drugs used.**

***Keywords*—Machine learning; Drug; Expert system; Forward chaining; Certainty factor**

I. INTRODUCTION

Social environment factors have influenced others to engage in drug use [1] [2]. Adult figures who are addicted to drugs have a great influence on the behavior of others to become addicted [3]. Poor, low skills, life pressure, anxiety, and deviant behavior are factors that also lead to drug use [3]. Relaxing, drinking, staying up late, increasing enthusiasm, and relieving stress are other triggering factors for many drug use among young people [4]. Drug use, including amphetamines, marijuana, cocaine, heroin, and the like, are a major public health problem in physiological symptoms, resulting in behavioral changes, cognitive problems, and mental health [5]. Drug use also affects the physiology and behavior of future generations [2].

Drug abuse causes physical dependence (addiction) or relapse to continuously consume [6], although it has resulted in physical and psychological problems [6]. The previous research show that drug users are very high [7] [1] and increasing globally [8]. Drug abuse has become a global epidemic that affects human behavior [9]. Because drug use is very detrimental to the physical and psychological harm of the user, it is not surprising that this research related to drugs has attracted attention for scientific research [10].

Another factor that is often considered in medical services is the cost factor and its inflexibility (Bevan & Patel, 2016). Processes that are done manually tend to cause delays in medical diagnosis [11]. Using an electronic system can speed up handling time, improve work efficiency, and save costs [11]. Using an electronic system allows lower errors and eliminates omissions in deciding the test results and achieving the results [12]. The success of curing drug abuse and dependence is still limited; this includes the lack of success in the early identification of at-risk populations, resulting in increased death rates due to overdose [13]. In other words, not paying attention to the early symptoms of consuming drugs will be disastrous and make people who are loved suffer the destructive effects of the substance [14].

Meanwhile, if it turns out to be able to identify it early, it can prevent harmful consequences in the future that are sure to occur [14]. It means that there is a need for breakthroughs in developing platforms that can identify and screen patients susceptible to addiction after using opioid drugs [13]. Therefore, this research aims to develop a machine learning that has expertise like an expert. The expert system created can identify and screen or diagnose early drug users and the types of drugs used by using the Certainty Factor and Forward Chaining methods. The certainty factor method measures the certainty of the type of drug used by the user or patient who conducts consultations. On the other hand, forward chaining plays a role in the flow of the reasoning process from beginning to end based on data mining of physical symptoms of drug users and the types of drugs used (collected or explored previously).

Medical data is helpful as the knowledge that helps make scientific decisions regarding drug use [15]. Electronic medical use based on doctor's notes is useful for an effective treatment medium [10]. In the meantime, data mining is capable of electronic checks based on the patient's medical record [10].

Besides, the machine learning methods are a technique that can be useful for finding correlations based on the case for prediction purposes [16]. Unfortunately, machine learning is still few in the medical field due to technical problems [17]. Therefore, the simple machine learning method built in this research by imitating (studying) human knowledge in analyzing the physical symptoms of drug users and then implementing it in predicting drug users and identifying drug types used by users. It means that the expert system in machine learning has intelligence like an expert in diagnosing users and the types of drugs used from the physical symptoms that arise from drug users. Taking into account that the current use of information and communication technology (ICT) is growing or expanding very quickly or booming [18]. Therefore, the embodiment of the machine learning system in this research is website-based. So, anyone (the public) can access it from anywhere and has flexibility because it can work on various devices and operating systems. A machine learning system in this research is helpful for early diagnosis without having to examine a narcotics laboratory and without a doctor or expert.

It is necessary to know the percentage of machine learning efficacy in identifying drug users and the type of drug used. It means that further testing to determine the actual percentage of machine learning efficacy still needs to be done. This study makes this happen by comparing the test results achieved by machine learning based on symptoms of drug users compared to the test results achieved from laboratory tests of drug users' urine in identifying drug users and the type of drug used. If machine learning has high efficacy, it can save time and cost of drug testing for suspects or drug users by using machine learning compared to testing drugs on urine or blood for suspects and drug users.

Some recent works related to this research:

- Zhongheng Zhang (2016) introduced the k-nearest neighbor (kNN) method as a simple machine learning method for modeling [17]. The similarity between the research in this article and the previous one is that they both use a simple approach to machine learning. While the difference is that the research uses the certainty factor method and forward chaining for machine learning, while previous research uses the kNN method for machine learning. Another difference is that the previous research was focus on predicting the class from the new dataset to the most similar class. In contrast, the research in this article focuses on machine learning to diagnose drug users and the types of drugs used.
- Anthony Anggrawan, Khasnur Hidjah, and Jihadil Qudsi S. (2017) implement intelligent application programs to detect kidney failure [19]. The previous research and the research in this article have similarities in developing web application programs with the PHP programming language and MySQL database. In addition, the last analysis used medical data on failure cases to diagnose new renal illness issues using CBR (Case-Based Reasoning method). In contrast, the articles in this study use the expertise of drug experts (specialists) as knowledge of the application system for early diagnosis of drug users and the types of drugs used by drug users using the Forward Chaining and Certainty Factor methods.
- Kurnia Muludi, Radix Suharjo, Admi Syarif, and Fitria Ramadhani (2018) identified tomato plant diseases [20]. This previous research and the research in this article both implements forward chaining and certainty factor methods. However, the difference in the last research is to build an expert system to identify plant diseases based on android [20], while the research in the article builds an expert system to identify users of drugs and the types of drugs used based on the website.
- Munaiseche, Kaparang, and Rompas (2018) built an expert system to assist doctors in diagnosing eye diseases [21]. In contrast to the research in this article, it is to build an expert system to diagnose drug users and the types of drugs used. Furthermore, this previous research used the forward chaining method, while the research in this article uses the forward chaining method and certainty factor. The similarities between the previous study and the research in this article are that both use PHP and MySQL in building an expert system.
- Ninive Von Greiff and Lisa Skogens (2021) investigates a drug user recovery program for drug addiction [22]. The research method is the interview or qualitative approach [22]. The similarity of the research in this article with previous studies is that they both study drug users. The difference is in previous studies examining the results of addiction recovery on drugs with the interview method. Meanwhile, the research in this article builds machine learning that has an intelligent system to detect drug users and the types of drugs used.

The latest related work identifies that the article in this study has a novelty that no previous researcher has researched. Other strength of this research is conducting a comparative test to determine the efficacy of machine learning or expert systems developed in identifying users and the types of drugs used by users that have not been studied before.

The systematics of writing this paper is as follows: the following sub-section discusses the research methodology, which includes research data and research methods used. The next subsection discusses the results and discussion of the research. Finally, the conclusions obtained from the study results and suggestions for further research are narrated in the Conclusions subsection.

II. Research Methodology

This study is a case study at the Indonesian National Narcotics Agency (*Badan Narkotika Nasional* or BNN) in Mataram, Indonesia. The number of drug users used as samples to test the expertise and accuracy of the machine learning built in this study was 30. The selection of data samples in this study was random. This research's development of machine learning expertise consists of stages: knowledge acquisition, expert system design (programming), machine learning/expert system testing, and accuracy test (see Fig.1).

A. Knowledge Acquisition

For the system development stage, the effort made is to obtain knowledge from drug experts, which is used as a knowledge base to build the machine learning or expert systems. The method used in obtaining knowledge related to narcotics is the interview method. The knowledge gained is the

knowledge about the types of drugs and their symptoms. Based on the knowledge obtained, there are ten types of drugs that drug users dominantly use, and there are 27 types of drug symptoms

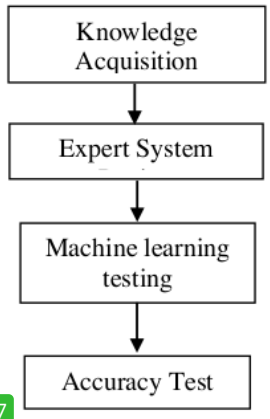

Fig.1. Stages in the development of an expert system on machine learning

In the knowledge acquisition stage, an expert from the Indonesian National Narcotics Agency (in Mataram, Indonesia) shares knowledge about drug use, including those related to symptoms and the types of drugs used by drug users. The knowledge gained at this stage serves as a knowledge base in building expertise from machine learning.

B. *Expert System Design*

The expert system design stage is a process for modeling the data that has been collected and designing an application system that is planned according to programming problems and the acquisition of knowledge obtained. This stage in computer science is known as planning the use case diagram design, data flow diagram (DFD) design, database design, and flowchart to be built on the application program. The programming stage is the implementation stage of the system design plan into a computer programming language. This research uses the PHP and MySQL programming languages. The computer application program that is built is a cloud application program. Application development with PHP programming language and MYSQL database makes application programs can run via the web. By being stored on the server computer, the application program is ubiquitous. The ubiquitous application program means that the application program can be accessed from anywhere and at any time [23].

C. *Machine learning testing*

The machine learning testing phase is the functional testing phase of the built application or black-box testing. Black-box testing is a test that no longer involves programming code or programming languages. In short, black-box testing on the expert system in this study is to determine whether the expert system built is under the list of desired system requirements.

D. *Accuracy Test*

Testing accuracy in machine learning is to determine the level of expertise of the expert system built in this study

III. Result and Discussion

A. *Knowledge Acquisition*

The knowledge acquisition stage is the stage of acquiring the required knowledge data. The acquired knowledge acquisition data is useful in solving programming logic in diagnosing users and the types of drugs used by drug users. Table I shows the code for the type of drug used by drug users. Meanwhile, Table II presents the code of symptoms caused by drug users.

TABLE I. List of Types of Drugs

No.	Drug Type Code	Drug Name
1	P01	Cocaine
2	P02	Marijuana
3	P03	Ecstasy
4	P04	Heroin
5	P05	Methamphetamine
6	P06	Hallucinogen
7	P07	Amphetamine
8	P08	Pethidine
9	P09	Codeine
10	P10	Morphine

TABLE II. List of Symptom of Drug User

Symptoms of Drug User			
Code	**Symptom**	**Code**	**Symptom**
G01	Out of breath	G15	Difficult to focus
G02	Anxious and restless	G16	Difficult to rest
G03	Nausea and vomiting	G17	Weight loss
G04	Diarrhea	G18	Dry mouth
G05	Convulsions	G19	Blurred vision
G06	Easy to get angry	G20	Changes in skin color
G07	Depression	G21	Constipation
G08	Changes in sleep patterns	G22	Stomachache
G09	Sweating	G23	Drowsiness
G10	Chills (Hot cold)	G24	Itching
G11	Shaking	G25	Difficulty urinating
G12	Insomnia	G26	Mood swings
G13	Fast heart rate	G27	Dizziness
G14	Increased blood pressure		

TABLE III. Rule Base of Types and Symptoms of Basic Diagnosis

Rule Base	
Drug Type	**Symptom**
P01	G01 and G02 and G03 and G04 and G05
P02	G06 and G02 and G07 and G08 and G09 and G10
P03	G05 and G11 and G12 and G13 and G14
P04	G15 and G02 and G07 and G16
P05	G11 and G01 and G16 and G17 and G18
P06	G09 and G11 and G18 and G19 and G10 and G14
P07	G18 and G03 and G04 and G05 and G01 and G20
P08	G07 and G13 and G05 and G03
P09	G27 and G03 and G18 and G21 and G22
P10	G023 and G24 and G09 and G25 and G26

After modeling the acquired knowledge acquisition data or knowledge representation (as shown in Table III) the next step is to implement it into the certainty factor algorithm. The certainty factor uses a value between 0.2 and 1.0 to assume a level of confidence in the data. A simulation of the calculation of the certainty factor was carried out based on the weight of symptoms arising from the type of drug used by drug users with weights of 0.8 and 1.0 according to the opinion of the drug expert (See Table IV).

TABLE IV. DETERMINATION OF DRUG SYMPTOM WEIGHT SCORE ACCORDING TO THE DECISION OF DRUG EXPERTS

No	Symptom	Weight Score
1	Very often	1
2	Often	0,8
3	Never	0

So, on the drug symptom weighted score given to the certainty factor, a score of 0 indicates that drug users do not experience these symptoms. If a drug user experiences symptoms, then the weighted score given for the frequently experienced symptoms is 0.8 and the most frequently experienced is 1.0, according to the drug expert's decision.

TABLE V. KNOWLEDGE BASE

	J-001	J-002	J-003	J-004	J-005	J-006	J-007	J-008	J-009	J-010
Symptom	CF	CF	CF	CF	CF	CF	CF	CF	CF	CF
Out of breath					0,8		1			
Anxious and restless	0,8	0,8		1						
Nausea and vomiting	0,8						0,8	0,8	0,8	
Diarrhea	1						0,8			
Convulsions	0,8		0,8				0,8	0,8		
Easy to get angry		1								
Depression		0,8		0,8				1		
Changes in sleep patterns		0,8		0,8						
Sweating		0,8				0,8				0,8
Chills		0,8				0,8				
Shaking			0,8		0,8	0,8				
Insomnia			0,8							
Fast heart rate			0,8					0,8		
Increased blood pressure			0,8			1				
Difficult to focus				1						
Difficult to rest					0,8					
Weight loss					0,8					
Dry mouth					1	0,8	0,8		1	
Blurred vision						0,8				
Changes in skin color							0,8			
Constipation									0,8	
Stomachache									0,8	
Drowsiness										1
Itching										0,8
Difficulty urinating										0,8
Mood swings										1
Dizziness									0,8	

This study's knowledge base of machine learning expertise is the symptoms, types of drugs, and CF rules obtained from drug experts (see Table V). The knowledge base is an essential component that contains the knowledge possessed by competent experts in the related field (i.e., narcotics in this study). Furthermore, the knowledge base is the basis for decision-making in an expert system, where this decision-making is related to the process of retrieving previously collected and stored knowledge.

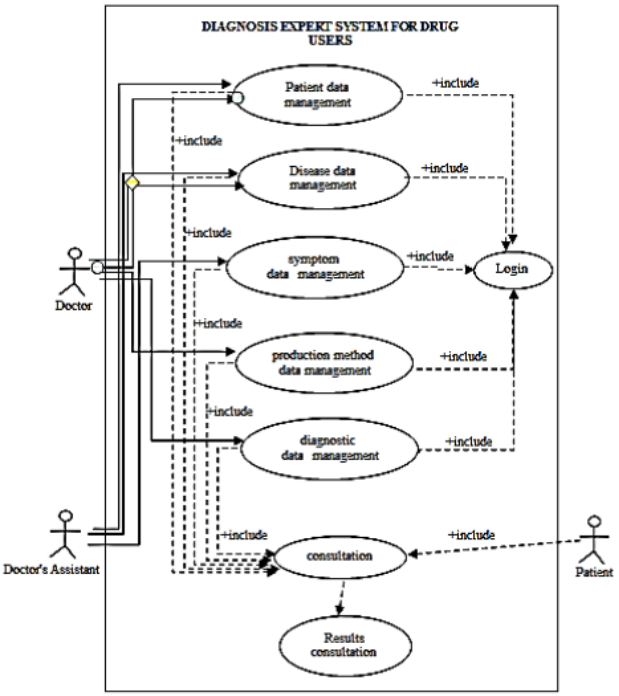

Fig. 2. Use Case Diagram on machine learning

This study has a database that stores records of users, patients, symptoms, and types of drugs, including diagnostic data, so it is necessary to design a data workflow model to realize a structured program.

The Data Flow Diagram (DFD) in Fig. 3 and Fig. 4 illustrate where the data flow comes from and where the data processing on the expert system is built. The context diagram in Fig. 3 shows the data flow of the system globally. In contrast, the overview diagram in Figure 4 shows a more detailed data flow that the system performs and its engagement with external data.

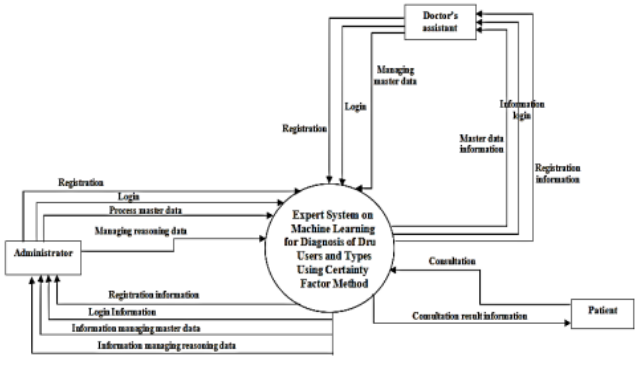

Fig. 3. Context Diagram of Data Flow on Machine Learning

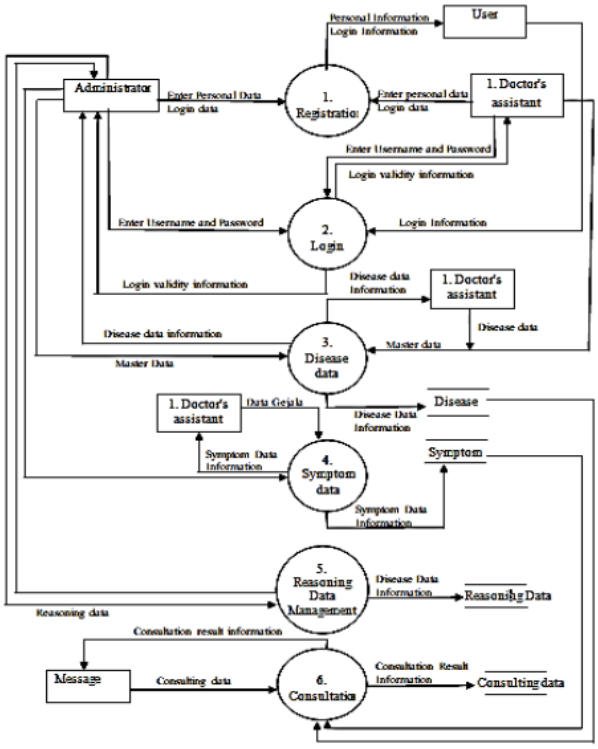

Fig. 4. Overview Diagram of Data Flow on Machine Learning

The flow diagram in fig. 5 shows a series of flow relationships in the expert system built in this study or shows the overall process sequence in building an expert system in this study.

The flow diagram contains a more detailed description of how each step of the procedure is actually carried out in building an expert system on machine learning that can diagnose users and the types of drugs used by users.

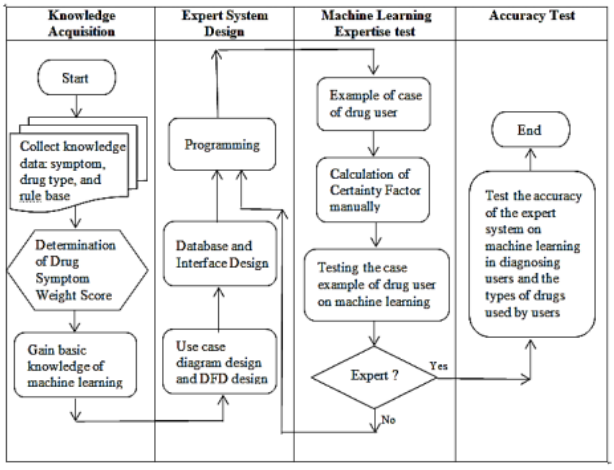

Fig. 5. Flow diagram of the whole Process of Building an Expert System on the Machine Learning

B. Machine learning testing

Expertise testing of machine learning is carried out using case samples from a patient. For example, in one case, a drug patient had symptoms of shortness of breath, depression, chills, anxiety and restlessness, and irritability. Symptoms of drug patients who have symptoms of shortness of breath, depression, chills, anxiety and restlessness, and irritability are symptoms of drug users: (1) Cocaine, (2) Cannabis, (3) Heroin, and (4) Amphetamine.

The formula for CF is:

CF[H,E] = CF[H] * CF[E]

CF Combine CF[H,E]1 = CF[H,E]1 + CF[H,E] 2 * (1 - CF [H,E] 1)

CF Combine CF[H,E] old3 = CF[H,E] old + CF[H, E] * (1 - CF[H,E] old)

Based on manual calculations, the results are as follows:

1. For J-001 = Cocaine

G01 = Out of breath (0.8)

CF[H,E] = CF[H] * CF[E]

= (0.8 * 0.8)

= 0.64

G02 = Anxious and restless (0.8)

CF[H,E] = CF[H] * CF[E]

= (0.8 * 0.8)

= 0.64

CFk1 = CF[H,E] 1 + CF[H,E]2 * (1 - CF[H,E]1)

= 0.64 + 0.64 * (1 - 0.64)

= 0.870

So the expert CF from the symptoms entered by the user for the type of drug Cocaine is probably 0.870 or 87%.

2. For J-002 = Marijuana

G07 = Depression (0,8)

CF[H,E] = CF[H] * CF[E]

= (0.8 * 0.8)

= 0.64

G10 = Chills (0.8)

CF[H,E] = CF[H] * CF[E]

= (0.8 * 0.8)

= 0.64

G02 = Anxious and restless (0.8)

CF[H,E] = CF[H] * CF[E]

= (0.8 * 0.8)

= 0.64

G06 = Easy to get angry (1)

CF[H,E] = CF[H] * CF[E]

= (0 * 1)

= 0

CFk1 = CF[H,E]1 + CF[H,E]2 * (1 - CF[H,E]1)

= 0.64 + 0.64 * (1 - 0.64)

= 0.870

CFk2 = CFk1 + CF[H,E]3 * (1 - CFk1)

= 0.870 + 0.64 * (1 - 0.870)

= 0.953

CFk3 = CFk2 + CF[H,E]4 * (1 – CFk2)

= 0.953 + 0 * (1 - 0.953)

= 0.953

So the CF of the symptoms entered by the user for the type of marijuana drug is likely to be 0.953 or 95%.

3. For J-004 = Heroin

G07 = Depression (0.8)

CF[H,E] = CF[H] * CF[E]

= (0.8 * 0.8)

= 0.64

G02 = Anxious and restless (1)

CF[H,E] = CF[H] * CF[E]

= (0 * 1)

= 0

CFk1 = CF[H,E]1 + CF[H,E]2 * (1 - CF[H,E]1)

= 0.64 + 0 * (1 - 0.64)

= 0.64

So the CF of the symptoms entered by the user for the type of heroin drug is most likely 0.64 or 64%.

4. For J-007 = Amphetamine

G01 = Out of breath (1)

CF[H,E]1 = CF[H]1 * CF[E]2

= (0.8 * 1)

= 0.8

So the CF of the symptoms entered by the user for the type of Amphetamine is most likely 0.8 or 80%.

Based on the value of manual calculations, the largest CF value is taken, which is 0.953 or 95% with the type of marijuana drug. It means the patient is using a type of marijuana drug. A case example is tested on an expert system application program (or on machine learning). If the patient's symptoms in the case sample (with the same symptoms) are entered into the expert system built in this study, the result of the process is as shown in Fig. 6.

HASIL KONSULTASI

Nama	:	Andi
Umur	:	24
Jenis Kelamin	:	Laki-laki
Pekerjaan	:	PNS
Alamat	:	ampenan

No	Pertanyaan	Jawaban
1	Sesak Nafas	TIDAK
2	Cemas dan Gelisah	IYA
3	Mual dan Muntah	TIDAK
4	Diare	TIDAK
5	Kejang-kejang	TIDAK
6	Mudah Marah	TIDAK
7	Depresi	IYA

Fig. 6. Screenshot of expert system questions about drug symptoms experienced by patients

Expert system testing on machine learning shows that the expert system has succeeded in correctly identifying the user and the type of drug used by the user. In order to know how accurate the machine learning expertise is, this study also tested several other patients by comparing the results with the urine test results at the Indonesian National Narcotic Agency laboratory in Mataram, Indonesia.

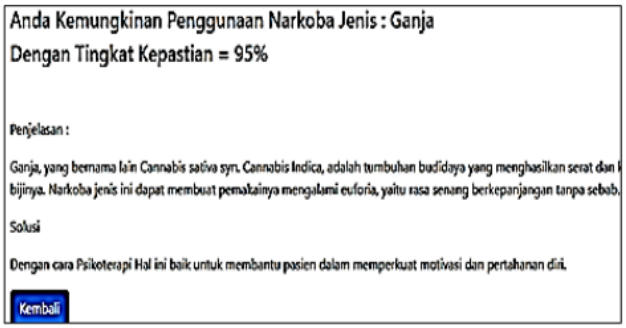

Fig.7. Screenshot of the expert system test results on the type of drug used by the patient

Fig. 7 describes it as follows: You are probably using a type of drug with a 95% certainty. Another narrative in figure 7 is: another name for cannabis is cannabis sativa. Cannabis sativa is a cultivated plant that contains fiber and narcotic substances in its seeds. This drug makes user experience euphoria, namely a prolonged feeling of pleasure for no reason. The cure is psychotherapy, which helps the patient strengthen the motivation to stop using it.

Expert system testing on machine learning shows that the expert system has succeeded in correctly identifying the user and the type of drug used by the user. In order to know how accurate the machine learning expertise is, this study also tested several other patients by comparing the results with the urine test results at the Indonesian National Narcotic Agency laboratory.

C. *Accuracy Test of Machine Learning*

The accuracy test of machine learning in this study is to determine the expert performance of the application system built-in diagnosing users and the types of drugs used. Testing the level of accuracy of machine learning expertise is to compare the suitability of the results with the urine test results from patients at the Indonesian National Narcotics Agency. In a trial of 30 times on 30 patients, there were 30 results of machine learning tests that can correctly identify drug users and as many as 24 machine learning test results that can detect the types of drugs used by drug users. It means that the test results on the data of 30 drug patients show that an expert system on machine learning built using the Certainty Factor method has expertise in diagnosing drug users up to 80 percent. The accuracy rate of up to 80 percent is obtained from the results of the calculation of 24 divided by 30 and multiplied by 100%. Table VI shows the details of the test results in diagnosing the type of abuse of drug users.

TABLE VI. Machine Learning Expertise Accuracy Test Result

Number	Case	System Result	Expert Result	Suitability
1	G01, G02, G03, G04, G05	Cocaine	Cocaine	Suitable
2	G06, G02, G07, G08, G09, G10	Marijuana	Marijuana	Suitable
3	G05, G11, G12, G13, G14	Ecstasy	Ecstasy	Suitable
4	G15, G02, G07, G16	Heroin	Heroin	Suitable
5	G11, G01, G16, G17, G18	Methamphetamine	Methamphetamine	Suitable
6	G09, G11, G18, G19, G10, G14	Hallucinogen	Hallucinogen	Suitable
7	G18, G03, G04, G05, G01, G20	Amphetamines	Amphetamines	Suitable
8	G07, G13, G05, G03	Pethidine	Pethidine	Suitable
9	G27, G03, G18, G21, G22	Codeine	Codeine	Suitable
10	G23, G24, G09, G25, G26	Morphine	Morphine	Suitable
11	G03, G06, G07, G15	Pethidine	Pethidine	Suitable
12	G01, G08, G09, G18	Codeine	Codeine	Suitable
13	G01, G02, G05, G09, G11,G15, G18	Codeine	Hallucinogen	Not suitable
14	G04, G12, G13, G17	Ecstasy	Ecstasy	Suitable
15	G08, G11, G17, G18, G19	Methamphetamine	Methamphetamine	Suitable
16	G01, G13, G15, G20, G22	Amphetamines	Amphetamines	Suitable
17	G17, G20, G21, G22, G23, G25, G27	Codeine	Morphine	Not suitable
18	G02, G07, G11, G15, G16	Pethidine	Heroin	Not suitable
19	G03, G07, G10, G14, G19	Pethidine	Pethidine	Suitable
20	G08, G11, G15, G18, G19, G20, G22, G23	Codeine	Amphetamines	Not suitable
21	G02, G06, G07, G08, G09, G10, G19, G27	Marijuana	Marijuana	Suitable
22	G05, G11, G12, G14, G19, G20	Hallucinogen	Ecstasy	Not suitable
23	G01, G06, G07, G10, G24	Pethidine	Marijuana	Not suitable
24	G03, G07, G09, G21	Morphine	Morphine	Suitable
25	G05, G12, G17, G22	Ecstasy	Ecstasy	Suitable
26	G07, G12, G15, G26	Heroin	Heroin	Suitable
27	G02, G08, G15, G23	Heroin	Heroin	Suitable
28	G06, G09, G15, G26	Morphine	Morphine	Suitable
29	G07, G11, G18, G23, G27	Codeine	Codeine	Suitable
30	G08, G15, G25, G26	Morphine	Morphine	Suitable

IV. CONCLUSION

The results of this study found that: (a). Machine learning in this study can predict drug users and types of drugs based on the symptoms that drug users complain about. (b). This study machine learning acquired knowledge about the symptoms of drug users, types of drugs, and basic knowledge related to the weight of the certainty factor of each type of drug and the symptoms caused so that it can diagnose drug users and the types of drugs used by users. (c). The accuracy of machine learning achieved in this study in predicting the types of drugs used by users and the types of drugs used by users reached 80%. (d). The expert system in this research is website-based, so that the expert system from this research can be used by various parties and in different places to identify users and the types of drugs used by users.

The implication of this research result is that the expert system built in this study can be a tool (choice) to replace or complete the testing system for drug users through urine testing in the laboratory.

The drawback of the results of this study is that machine learning expertise in this study is only limited to simple machine learning, as is the case with simple machine learning which was built on previous research by Zhongheng Zhang (2016), which used the KNN method in building learning machines. Furthermore, the machine learning expertise generated from this research is only limited to the expertise possessed in accordance with the knowledge obtained (symptoms, types of drug abuse, rule base, and calculation of certainty factor) under study. Therefore, further research needs to build machine learning that can increase its expertise based on more new data and use another method.

REFERENCES

[1] R. Jiménez, J. Anupol, B. Cajal, and E. Gervilla, "Data mining techniques drug use research," *Addict. Behav. Reports*, vol. 8, pp. 128–135, 2018.

[2] F. M. Vassoler, E. M. Byrnes, and R. C. Pierce, "The impact of exposure to addictive drugs on future generations: Physiological and behavioral effects," *Neuropharmacology*, vol. 76, no. PART B, pp. 269–275, 2014.

[3] P. K. Shanmugam, "The Influence of Social Factors in Drug Addiction–A Mini Review of Work by Miller & Carroll (2006)," *J. Alcohol. Drug end.*, vol. 05, no. 04, pp. 4–6, 2017.

[4] Boys, J. Marsden, and J. Strang, "Understanding reasons for drug use amongst young people: A functional perspective," *Health Educ. Res.*, vol. 16, no. 4, pp. 457–469, 2001.

[5] G. López, L. M. Orchowski, M. K. Reddy, J. Nargiso, and J. E. Johnson, "A review of research-supported group treatments for drug use rders," *BMC Public Health*, vol. 16, no. 51, pp. 1–21, 2021.

[6] Z. Justinova, L. V. Panlilio, and S. R. Goldberg, "Drug Addiction," *Natl. Libr. Medicine*, vol. 1, no. 1, pp. 310–335, 2009.

[7] T. Saah, "The evolutionary origins and significance of drug addiction," *Harm Reduct. J.*, vol. 2, pp. 1–7, 2005.

[8] G. Leshner M.Stevens, S. Kim, N. Kim, T. L.Wagener, and A. C. Villantie, "Cognitive and affective responses to marijuana prevention and educational messaging," *Drug Alcohol Depend.*, vol. 225, no. August, pp. 1–3, 2021.

[9] M. Zaman *et al.*, "Drug abuse among the students," *Pakistan J. Pharm.* , vol. 1, no. 1, p. 41, 2015.

[10] L. W. Chou, K. M. Chang, and I. Puspitasari, "Drug Abuse Research

Trend Investigation with Text Mining," *Comput. Math. Methods Med.*, vol. 2020, pp. 1–8, 2020.

[11] A. Bevan and N. Patel, "An Electronic Prescription Alerting System-Improving the Discharge Medicines Process," *Arch. Dis. Child.*, vol. 101, no. 9, p. e2.55-e2, 2016.

[12] A. Tsyben, N. Gooding, and W. Kelsall, "Assessing the Impact of a Newly Introduced Electronic Prescribing System Across a Paediatric Department – Lessons Learned," *Arch. Dis. Child.*, vol. 101, no. 9, p. 3-e2, 2016.

[13] M. Mahmoudi, S. Pakpour, and G. Perry, "Drug-Abuse Nanotechnology: Opportunities and Challenges," *ACS Chem. Neurosci.*, vol. 9, no. 10, pp. 2288–2298, 2018.

[14] J. Redman, "Recognizing the Warning Signs of Drug Addiction : What You Need to Know," *Mental Health and Counseling Studies*. pp. 1–7, 2021.

[15] N. Jojen, "A Survey Paper on Data Mining Techniques in Drug Industry," *. J. Eng. Res. Technol.*, vol. 3, no. 30, pp. 296–299, 2015.

[16] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or n category," *Front. Chem.*, vol. 6, no. May, pp. 1–11, 2018.

[17] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. nsl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.

[18] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 4, pp. 319–326, 2021.

[19] A. Anggrawan, K. Hidjah, and Q. S. Jihadil, "Kidney failure diagnosis based on case-based reasoning (CBR) method and statistical analysis," in *2016 International Conference on Informatics and Computing, ICIC 16*, 2017, pp. 298–303.

[20] K. Muludi, R. Suharjo, A. Syarif, and F. Ramadhani, "Implementation of forward chaining and certainty factor m d on android-based expert stem of tomato diseases identification," *Int. J. Adv. Comput. Sci. Appl.*, l. 9, no. 9, pp. 451–456, 2018.

[21] C. P. C. Munaiseche, D. R. Kaparang, and P. T. D. Rompas, "An Expert System for Diagnosing Eye Diseases using Forward Chaining Method," in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. no. 1, pp. 1–8.

[22] N. Von Greiff and L. Skogens, "Recovery and identity: a five-year follow-up of persons treated in 12-step-related programs," *Drugs Educ. ev. Policy*, pp. 1–10, 2021.

[23] A. Anggrawan, N. Ibrahim, S. Muslim, and C. Satria, "Interaction between Learning Style and Gender in Mixed Learning with 40 % Face-to-face Learning and 60 % Online Learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, pp. 407–413, 2019.

Similarity Check _ Machine Learning for Diagnosing Drug Users and Types of Drugs used

ORIGINALITY REPORT

15%
SIMILARITY INDEX

PRIMARY SOURCES

1	repository.universitasbumigora.ac.id Internet	244 words — 5%
2	www.jatit.org Internet	55 words — 1%
3	Mohammed Jasim A. Alkhafaji, Abbas Fadhil Aljuboori, Abdullahi Abdu Ibrahim. "Clean medical data and predict heart disease", 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2020 Crossref	47 words — 1%
4	journal.uad.ac.id Internet	39 words — 1%
5	arxiv.org Internet	36 words — 1%
6	onlinelibrary.wiley.com Internet	32 words — 1%
7	www.ijrte.org Internet	30 words — 1%
8	Joan Angelina Widians, Novianti Puspitasari, Aidil Febriansyah. "Disease Diagnosis System Using	26 words — < 1%

Certainty Factor", 2019 International Conference on Electrical, Electronics and Information Engineering (ICEEIE), 2019
Crossref

9	turcomat.org Internet	26 words — < 1%
10	Arifin Ika Nugroho, Pulung Riyanto, Jori Lahinda, Serli Hatul Hidayat. "Effectiveness of physical condition training on the fitness of drug abuse victims, BRSPP Yogyakarta", Enfermería Clínica, 2020 Crossref	20 words — < 1%
11	www.ijeat.org Internet	20 words — < 1%
12	academic.oup.com Internet	19 words — < 1%
13	www.tandfonline.com Internet	18 words — < 1%
14	Anthony Anggrawan, Nurdin Ibrahim, Suyitno Muslim, Christofer Satria. "Interaction between Learning Style and Gender in Mixed Learning with 40% Face-to-face Learning and 60% Online Learning", International Journal of Advanced Computer Science and Applications, 2019 Crossref	17 words — < 1%
15	publikasi.mercubuana.ac.id Internet	16 words — < 1%
16	Ashok Singh. "Novel nanoparticle-based treatment approaches", Elsevier BV, 2021 Crossref	15 words — < 1%
17	Edgar A. Whitley. "Two approaches to developing expert systems: A consideration of formal and	15 words — < 1%

semi-formal domains", AI & Society, 1991
Crossref

18 Ania Wajnberg, Katherine Ornstein, Meng Zhang, Kristofer L. Smith, Theresa Soriano. "Symptom Burden in Chronically Ill Homebound Individuals", Journal of the American Geriatrics Society, 2013 — 14 words — < 1%
Crossref

19 W Uriawan, K Manaf, U Syaripudin, M Mujiburrahman. "Expert system for diagnosing narcissistic personality disorders using certainty factor and forward chaining methods", IOP Conference Series: Materials Science and Engineering, 2021 — 14 words — < 1%
Crossref

20 Win Min Thit, Sai Wai Yan Myint Thu, Jaranit Kaewkungwal, Ngamphol Soonthornworasiri et al. "User Acceptance of Electronic Medical Record System: Implementation at Marie Stopes International, Myanmar", Healthcare Informatics Research, 2020 — 13 words — < 1%
Crossref

21 repositorio.unal.edu.co — 12 words — < 1%
Internet

22 Wu, L.T.. "Hallucinogen-related disorders in a national sample of adolescents: The influence of ecstasy/MDMA use", Drug and Alcohol Dependence, 20090901 — 11 words — < 1%
Crossref

23 Hengki Tamando Sihotang, Fristi Riandari, R.Mahdalena Simanjorang, Agustina Simangunsong, Penda Sudarto Hasugian. "Expert System for Diagnosis Chicken Disease using Bayes Theorem", Journal of Physics: Conference Series, 2019 — 10 words — < 1%
Crossref

24	Zhang, Rui-san, Zhen He, Wei-dong Jin, and Rui Wang. "Effects of the cannabinoid 1 receptor peptide ligands hemopressin, (m)RVD-hemopressin(?) and (m)VD-hemopressin(?) on memory in novel object and object location recognition tasks in normal young and A?1-42-treated mice", Neurobiology of Learning and Memory, 2016. Crossref	10 words —	< 1%
25	Li Ju, Aihua Lyu, Hongxia Hao, Wen Shen, Hua Cui. "Deep Learning-Assisted Three-Dimensional Fluorescence Difference Spectroscopy for Identification and Semiquantification of Illicit Drugs in Biofluids", Analytical Chemistry, 2019 Crossref	9 words —	< 1%
26	Glenn Leshner, Elise M. Stevens, Amy M. Cohn, Seunghyun Kim, Narae Kim, Theodore L. Wagener, Andrea C. Villanti. "Cognitive and affective responses to marijuana prevention and educational messaging", Drug and Alcohol Dependence, 2021 Crossref	8 words —	< 1%
27	Sutrisno Arianto Pasaribu, Poltak Sihombing, Suherman Suherman. "Expert System for Diagnosing Dental and Mouth Diseases with a Website-Based Certainty Factor (CF) Method", 2020 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT), 2020 Crossref	8 words —	< 1%
28	ojs.trigunadharma.ac.id Internet	7 words —	< 1%
29	Deny Adhar, Syafrizal, Erwin Panggabean, Lahmudin Sipahutar, Muhammad Reza Fahlevi. "Expert System to Diagnose Gastritis Disease Using the	6 words —	< 1%

Certainty Factor Method", 2019 7th International Conference on Cyber and IT Service Management (CITSM), 2019
Crossref

30 Ninive von Greiff, Lisa Skogens. "Recovery and identity: a five-year follow-up of persons treated in 12-step-related programs", Drugs: Education, Prevention and Policy, 2021 — 6 words — < 1%
Crossref

EXCLUDE QUOTES	OFF	EXCLUDE MATCHES	OFF
EXCLUDE BIBLIOGRAPHY	OFF		

Paper Submission System

Submission Type *
- (●) Research Paper
- ( ) Review/ Survey Paper

Paper Title *
Machine Learning for Diagnosing Drug Users and Types of Drugs u

Abstract *
machine learning developed can be used to diagnose drug users and distinguish the types of drugs used with an accuracy of up to 80%. The implications of the expert system on machine learning are an alternative method for narcotics officers and medical doctors in diagnosing drug users and the types of drugs used.

Upload Paper *
Choose File ijacsa_Machi…gs used.pdf
Max file size: 5 MB. Allowed file types: docx, pdf.

Primary Domain *
Machine Learning

Secondary Domain *
Data Mining

Country *
Indonesia

Author Details

Please add all Author Names and Email IDs comma separated.

Author Name(s) *
Anthony Anggrawan, Christofer Satria, Che Ku Nuraini, Lusiar

Author Email ID(s) *
anthony.anggrawan@universitasbumigora.ac.id, chris@univer

How did you hear about this journal *
University

Agreement to Publication Ethics *
[x] I confirm, acknowledge and agree that the submission is original and not submitted or published elsewhere, and if the authors have used the work and/or words of others, it has been appropriately cited or referenced.

Submit

Similarity Check _ Machine Learning for Diagnosing Drug Users and Types of Drugs used

By Anthony Anggrawan

WORD COUNT	5295	TIME SUBMITTED	17-OCT-2021 07:47PM
		PAPER ID	77793429

Machine Learning for Diagnosing Drug Users and Types of Drugs used

Anthony Anggrawan[1]
Ni Gusti Ayu Dasriani[5]
Mayadi[6]
Computer Science Study Program
Bumigora University, Mataram, Indonesia

Christofer Satria[2]
Visual Communication Design Study Program
Bumigora University, Mataram, Indonesia

Che Ku Nuraini[3]
Department of Centre of Research and Innovation Management
Universiti Teknikal Malaysia Melaka, Malaysia

Lusiana[4]
Technical Information Study Program
STMIK AMIK Riau, Riau, Indonesia

***Abstract*— Drug use is very detrimental to the physical and psychological health of users. Drug abuse also causes addiction and is a global epidemic. Therefore it is not surprising that scientific research related to drugs has attracted attention for research. However, many factors become obstacles in the medical services of the drug user, including cost, flexibility, and a slow process. Meanwhile, electronic systems can speed up handling time, improve work efficiency, save costs and reduce inspection errors. It means that a breakthrough is needed in developing a platform that can identify drug users. Therefore, this research aims to build machine learning with expertise like an expert who can diagnose drug users and distinguish the types of drugs used by drug users. The expert system on machine learning was developed using the Forward Chaining and Certainty Factor methods. This study concludes that the expert system on machine learning developed can be used to diagnose drug users and distinguish the types of drugs used with an accuracy of up to 80%. The implications of the expert system on machine learning are an alternative method for narcotics officers and medical doctors in diagnosing drug users and the types of drugs used.**

***Keywords*—Machine learning; Drug; Expert system; Forward chaining; Certainty factor**

I. INTRODUCTION

Social environment factors have influenced others to engage in drug use [1] [2]. Adult figures who are addicted to drugs have a great influence on the behavior of others to become addicted [3]. Poor, low skills, life pressure, anxiety, and deviant behavior are factors that also lead to drug use [3]. Relaxing, drinking, staying up late, increasing enthusiasm, and relieving stress are other triggering factors for many drug use among young people [4]. Drug use, including amphetamines, marijuana, cocaine, heroin, and the like, are a major public health problem in physiological symptoms, resulting in behavioral changes, cognitive problems, and mental health [5]. Drug use also affects the physiology and behavior of future generations [2].

Drug abuse causes physical dependence (addiction) or relapse to continuously consume [6], although it has resulted in physical and psychological problems [6]. The previous research show that drug users are very high [7] [1] and increasing globally [8]. Drug abuse has become a global epidemic that affects human behavior [9]. Because drug use is very detrimental to the physical and psychological harm of the user, it is not surprising that this research related to drugs has attracted attention for scientific research [10].

Another factor that is often considered in medical services is the cost factor and its inflexibility (Bevan & Patel, 2016). Processes that are done manually tend to cause delays in medical diagnosis [11]. Using an electronic system can speed up handling time, improve work efficiency, and save costs [11]. Using an electronic system allows lower errors and eliminates omissions in deciding the test results and achieving the results [12]. The success of curing drug abuse and dependence is still limited; this includes the lack of success in the early identification of at-risk populations, resulting in increased death rates due to overdose [13]. In other words, not paying attention to the early symptoms of consuming drugs will be disastrous and make people who are loved suffer the destructive effects of the substance [14].

Meanwhile, if it turns out to be able to identify it early, it can prevent harmful consequences in the future that are sure to occur [14]. It means that there is a need for breakthroughs in developing platforms that can identify and screen patients susceptible to addiction after using opioid drugs [13]. Therefore, this research aims to develop a machine learning that has expertise like an expert. The expert system created can identify and screen or diagnose early drug users and the types of drugs used by using the Certainty Factor and Forward Chaining methods. The certainty factor method measures the certainty of the type of drug used by the user or patient who conducts consultations. On the other hand, forward chaining plays a role in the flow of the reasoning process from beginning to end based on data mining of physical symptoms of drug users and the types of drugs used (collected or explored previously).

Medical data is helpful as the knowledge that helps make scientific decisions regarding drug use [15]. Electronic medical use based on doctor's notes is useful for an effective treatment medium [10]. In the meantime, data mining is capable of electronic checks based on the patient's medical record [10].

Besides, the machine learning methods are a technique that can be useful for finding correlations based on the case for prediction purposes [16]. Unfortunately, machine learning is still few in the medical field due to technical problems [17]. Therefore, the simple machine learning method built in this research by imitating (studying) human knowledge in analyzing the physical symptoms of drug users and then implementing it in predicting drug users and identifying drug types used by users. It means that the expert system in machine learning has intelligence like an expert in diagnosing users and the types of drugs used from the physical symptoms that arise from drug users. Taking into account that the current use of information and communication technology (ICT) is growing or expanding very quickly or booming [18]. Therefore, the embodiment of the machine learning system in this research is website-based. So, anyone (the public) can access it from anywhere and has flexibility because it can work on various devices and operating systems. A machine learning system in this research is helpful for early diagnosis without having to examine a narcotics laboratory and without a doctor or expert.

It is necessary to know the percentage of machine learning efficacy in identifying drug users and the type of drug used. It means that further testing to determine the actual percentage of machine learning efficacy still needs to be done. This study makes this happen by comparing the test results achieved by machine learning based on symptoms of drug users compared to the test results achieved from laboratory tests of drug users' urine in identifying drug users and the type of drug used. If machine learning has high efficacy, it can save time and cost of drug testing for suspects or drug users by using machine learning compared to testing drugs on urine or blood for suspects and drug users.

Some recent works related to this research:

- Zhongheng Zhang (2016) introduced the k-nearest neighbor (kNN) method as a simple machine learning method for modeling [17]. The similarity between the research in this article and the previous one is that they both use a simple approach to machine learning. While the difference is that the research uses the certainty factor method and forward chaining for machine learning, while previous research uses the kNN method for machine learning. Another difference is that the previous research was focus on predicting the class from the new dataset to the most similar class. In contrast, the research in this article focuses on machine learning to diagnose drug users and the types of drugs used.
- Anthony Anggrawan, Khasnur Hidjah, and Jihadil Qudsi S. (2017) implement intelligent application programs to detect kidney failure [19]. The previous research and the research in this article have similarities in developing web application programs with the PHP programming language and MySQL database. In addition, the last analysis used medical data on failure cases to diagnose new renal illness issues using CBR (Case-Based Reasoning method). In contrast, the articles in this study use the expertise of drug experts (specialists) as knowledge of the application system for early diagnosis of drug users and the types of drugs used by drug users using the Forward Chaining and Certainty Factor methods.
- Kurnia Muludi, Radix Suharjo, Admi Syarif, and Fitria Ramadhani (2018) identified tomato plant diseases [20]. This previous research and the research in this article both implements forward chaining and certainty factor methods. However, the difference in the last research is to build an expert system to identify plant diseases based on android [20], while the research in the article builds an expert system to identify users of drugs and the types of drugs used based on the website.
- Munaiseche, Kaparang, and Rompas (2018) built an expert system to assist doctors in diagnosing eye diseases [21]. In contrast to the research in this article, it is to build an expert system to diagnose drug users and the types of drugs used. Furthermore, this previous research used the forward chaining method, while the research in this article uses the forward chaining method and certainty factor. The similarities between the previous study and the research in this article are that both use PHP and MySQL in building an expert system.
- Ninive Von Greiff and Lisa Skogens (2021) investigates a drug user recovery program for drug addiction [22]. The research method is the interview or qualitative approach [22]. The similarity of the research in this article with previous studies is that they both study drug users. The difference is in previous studies examining the results of addiction recovery on drugs with the interview method. Meanwhile, the research in this article builds machine learning that has an intelligent system to detect drug users and the types of drugs used.

The latest related work identifies that the article in this study has a novelty that no previous researcher has researched. Other strength of this research is conducting a comparative test to determine the efficacy of machine learning or expert systems developed in identifying users and the types of drugs used by users that have not been studied before.

The systematics of writing this paper is as follows: the following sub-section discusses the research methodology, which includes research data and research methods used. The next subsection discusses the results and discussion of the research. Finally, the conclusions obtained from the study results and suggestions for further research are narrated in the Conclusions subsection.

II. Research Methodology

This study is a case study at the Indonesian National Narcotics Agency (*Badan Narkotika Nasional* or BNN) in Mataram, Indonesia. The number of drug users used as samples to test the expertise and accuracy of the machine learning built in this study was 30. The selection of data samples in this study was random. This research's development of machine learning expertise consists of stages: knowledge acquisition, expert system design (programming), machine learning/expert system testing, and accuracy test (see Fig.1).

A. Knowledge Acquisition

For the system development stage, the effort made is to obtain knowledge from drug experts, which is used as a knowledge base to build the machine learning or expert systems. The method used in obtaining knowledge related to narcotics is the interview method. The knowledge gained is the

knowledge about the types of drugs and their symptoms. Based on the knowledge obtained, there are ten types of drugs that drug users dominantly use, and there are 27 types of drug symptoms

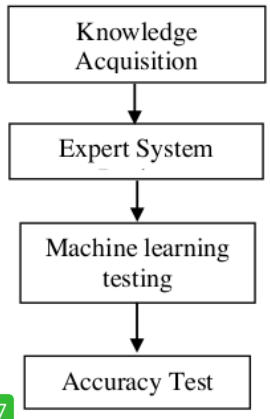

Fig.1. Stages in the development of an expert system on machine learning

In the knowledge acquisition stage, an expert from the Indonesian National Narcotics Agency (in Mataram, Indonesia) shares knowledge about drug use, including those related to symptoms and the types of drugs used by drug users. The knowledge gained at this stage serves as a knowledge base in building expertise from machine learning.

B. *Expert System Design*

The expert system design stage is a process for modeling the data that has been collected and designing an application system that is planned according to programming problems and the acquisition of knowledge obtained. This stage in computer science is known as planning the use case diagram design, data flow diagram (DFD) design, database design, and flowchart to be built on the application program. The programming stage is the implementation stage of the system design plan into a computer programming language. This research uses the PHP and MySQL programming languages. The computer application program that is built is a cloud application program. Application development with PHP programming language and MYSQL database makes application programs can run via the web. By being stored on the server computer, the application program is ubiquitous. The ubiquitous application program means that the application program can be accessed from anywhere and at any time [23].

C. *Machine learning testing*

The machine learning testing phase is the functional testing phase of the built application or black-box testing. Black-box testing is a test that no longer involves programming code or programming languages. In short, black-box testing on the expert system in this study is to determine whether the expert system built is under the list of desired system requirements.

D. *Accuracy Test*

Testing accuracy in machine learning is to determine the level of expertise of the expert system built in this study

III. Result and Discussion

A. *Knowledge Acquisition*

The knowledge acquisition stage is the stage of acquiring the required knowledge data. The acquired knowledge acquisition data is useful in solving programming logic in diagnosing users and the types of drugs used by drug users. Table I shows the code for the type of drug used by drug users. Meanwhile, Table II presents the code of symptoms caused by drug users.

TABLE I. List of Types of Drugs

No.	Drug Type Code	Drug Name
1	P01	Cocaine
2	P02	Marijuana
3	P03	Ecstasy
4	P04	Heroin
5	P05	Methamphetamine
6	P06	Hallucinogen
7	P07	Amphetamine
8	P08	Pethidine
9	P09	Codeine
10	P10	Morphine

TABLE II. List of Symptom of Drug User

Symptoms of Drug User			
Code	**Symptom**	**Code**	**Symptom**
G01	Out of breath	G15	Difficult to focus
G02	Anxious and restless	G16	Difficult to rest
G03	Nausea and vomiting	G17	Weight loss
G04	Diarrhea	G18	Dry mouth
G05	Convulsions	G19	Blurred vision
G06	Easy to get angry	G20	Changes in skin color
G07	Depression	G21	Constipation
G08	Changes in sleep patterns	G22	Stomachache
G09	Sweating	G23	Drowsiness
G10	Chills (Hot cold)	G24	Itching
G11	Shaking	G25	Difficulty urinating
G12	Insomnia	G26	Mood swings
G13	Fast heart rate	G27	Dizziness
G14	Increased blood pressure		

TABLE III. Rule Base of Types and Symptoms of Basic Diagnosis

Rule Base	
Drug Type	**Symptom**
P01	G01 and G02 and G03 and G04 and G05
P02	G06 and G02 and G07 and G08 and G09 and G10
P03	G05 and G11 and G12 and G13 and G14
P04	G15 and G02 and G07 and G16
P05	G11 and G01 and G16 and G17 and G18
P06	G09 and G11 and G18 and G19 and G10 and G14
P07	G18 and G03 and G04 and G05 and G01 and G20
P08	G07 and G13 and G05 and G03
P09	G27 and G03 and G18 and G21 and G22
P10	G023 and G24 and G09 and G25 and G26

After modeling the acquired knowledge acquisition data or knowledge representation (as shown in Table III) the next step is to implement it into the certainty factor algorithm. The certainty factor uses a value between 0.2 and 1.0 to assume a level of confidence in the data. A simulation of the calculation of the certainty factor was carried out based on the weight of symptoms arising from the type of drug used by drug users with weights of 0.8 and 1.0 according to the opinion of the drug expert (See Table IV).

TABLE IV. DETERMINATION OF DRUG SYMPTOM WEIGHT SCORE ACCORDING TO THE DECISION OF DRUG EXPERTS

No	Symptom	Weight Score
1	Very often	1
2	Often	0,8
3	Never	0

So, on the drug symptom weighted score given to the certainty factor, a score of 0 indicates that drug users do not experience these symptoms. If a drug user experiences symptoms, then the weighted score given for the frequently experienced symptoms is 0.8 and the most frequently experienced is 1.0, according to the drug expert's decision.

TABLE V. KNOWLEDGE BASE

	J-001	J-002	J-003	J-004	J-005	J-006	J-007	J-008	J-009	J-010
Symptom	CF	CF	CF	CF	CF	CF	CF	CF	CF	CF
Out of breath					0,8		1			
Anxious and restless	0,8	0,8		1						
Nausea and vomiting	0,8						0,8	0,8	0,8	
Diarrhea	1						0,8			
Convulsions	0,8		0,8				0,8	0,8		
Easy to get angry		1								
Depression		0,8		0,8				1		
Changes in sleep patterns		0,8		0,8						
Sweating		0,8				0,8				0,8
Chills		0,8				0,8				
Shaking			0,8		0,8	0,8				
Insomnia			0,8							
Fast heart rate			0,8					0,8		
Increased blood pressure			0,8			1				
Difficult to focus				1						
Difficult to rest					0,8					
Weight loss					0,8					
Dry mouth					1	0,8	0,8		1	
Blurred vision						0,8				
Changes in skin color							0,8			
Constipation									0,8	
Stomachache									0,8	
Drowsiness										1
Itching										0,8
Difficulty urinating										0,8
Mood swings										1
Dizziness									0,8	

This study's knowledge base of machine learning expertise is the symptoms, types of drugs, and CF rules obtained from drug experts (see Table V). The knowledge base is an essential component that contains the knowledge possessed by competent experts in the related field (i.e., narcotics in this study). Furthermore, the knowledge base is the basis for decision-making in an expert system, where this decision-making is related to the process of retrieving previously collected and stored knowledge.

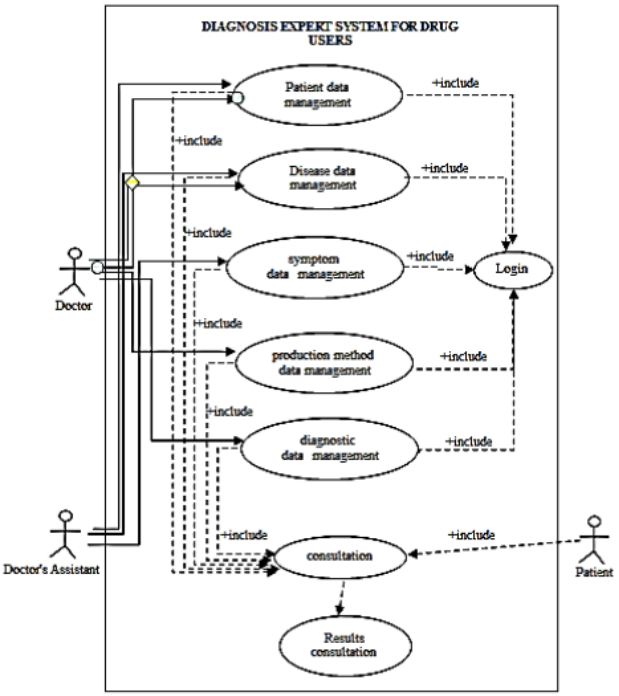

Fig. 2. Use Case Diagram on machine learning

This study has a database that stores records of users, patients, symptoms, and types of drugs, including diagnostic data, so it is necessary to design a data workflow model to realize a structured program.

The Data Flow Diagram (DFD) in Fig. 3 and Fig. 4 illustrate where the data flow comes from and where the data processing on the expert system is built. The context diagram in Fig. 3 shows the data flow of the system globally. In contrast, the overview diagram in Figure 4 shows a more detailed data flow that the system performs and its engagement with external data.

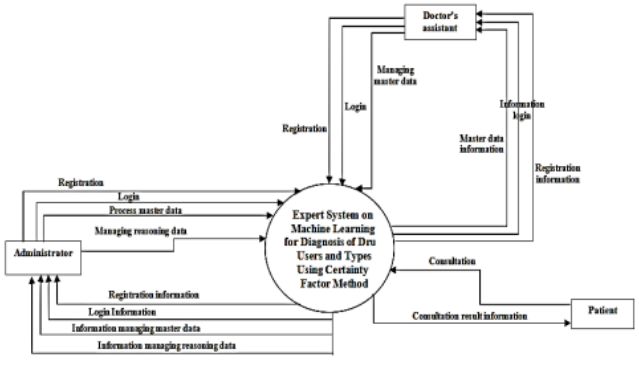

Fig. 3. Context Diagram of Data Flow on Machine Learning

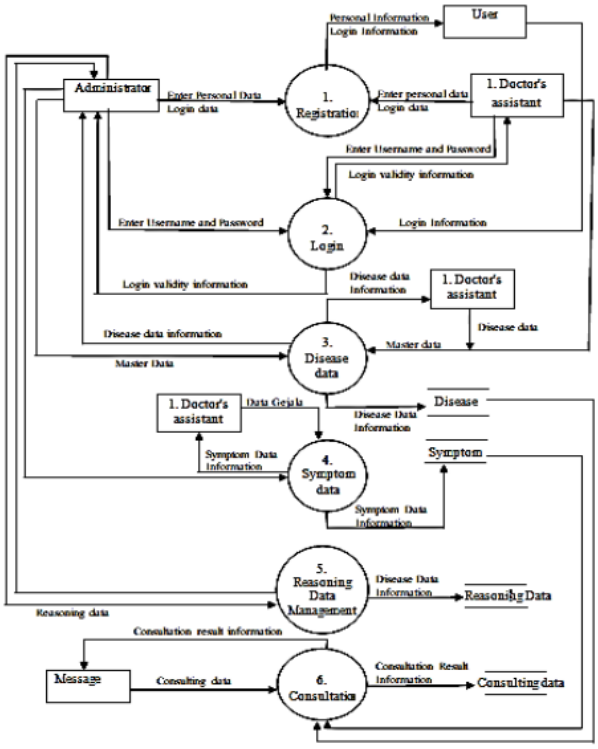

Fig. 4. Overview Diagram of Data Flow on Machine Learning

The flow diagram in fig. 5 shows a series of flow relationships in the expert system built in this study or shows the overall process sequence in building an expert system in this study.

The flow diagram contains a more detailed description of how each step of the procedure is actually carried out in building an expert system on machine learning that can diagnose users and the types of drugs used by users.

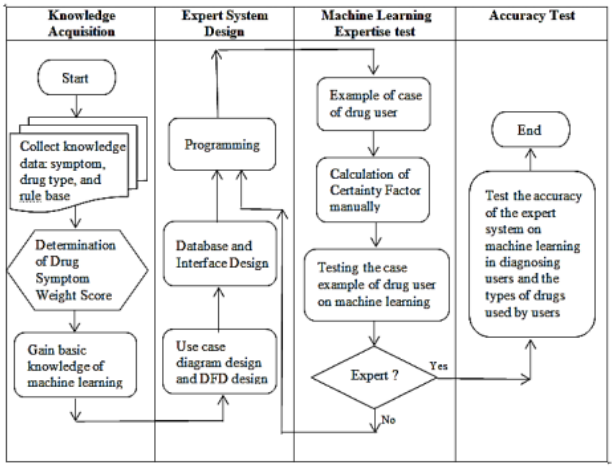

Fig. 5. Flow diagram of the whole Process of Building an Expert System on the Machine Learning

B. Machine learning testing

Expertise testing of machine learning is carried out using case samples from a patient. For example, in one case, a drug patient had symptoms of shortness of breath, depression, chills, anxiety and restlessness, and irritability. Symptoms of drug patients who have symptoms of shortness of breath, depression, chills, anxiety and restlessness, and irritability are symptoms of drug users: (1) Cocaine, (2) Cannabis, (3) Heroin, and (4) Amphetamine.

The formula for CF is:

CF[H,E] = CF[H] * CF[E]

CF Combine CF[H,E]1 = CF[H,E]1 + CF[H,E] 2 * (1 - CF [H,E] 1)

CF Combine CF[H,E] old3 = CF[H,E] old + CF[H, E] * (1 - CF[H,E] old)

Based on manual calculations, the results are as follows:

1. For J-001 = Cocaine

G01 = Out of breath (0.8)

CF[H,E] = CF[H] * CF[E]

= (0.8 * 0.8)

= 0.64

G02 = Anxious and restless (0.8)

CF[H,E] = CF[H] * CF[E]

= (0.8 * 0.8)

= 0.64

CFk1 = CF[H,E] 1 + CF[H,E]2 * (1 - CF[H,E]1)

= 0.64 + 0.64 * (1 - 0.64)

= 0.870

So the expert CF from the symptoms entered by the user for the type of drug Cocaine is probably 0.870 or 87%.

2. For J-002 = Marijuana

G07 = Depression (0,8)

CF[H,E] = CF[H] * CF[E]

= (0.8 * 0.8)

= 0.64

G10 = Chills (0.8)

CF[H,E] = CF[H] * CF[E]

= (0.8 * 0.8)

= 0.64

G02 = Anxious and restless (0.8)

CF[H,E] = CF[H] * CF[E]

= (0.8 * 0.8)

= 0.64

G06 = Easy to get angry (1)

CF[H,E] = CF[H] * CF[E]

= (0 * 1)

= 0

CFk1 = CF[H,E]1 + CF[H,E]2 * (1 - CF[H,E]1)

= 0.64 + 0.64 * (1 - 0.64)

= 0.870

CFk2 = CFk1 + CF[H,E]3 * (1 - CFk1)

= 0.870 + 0.64 * (1 - 0.870)

= 0.953

CFk3 = CFk2 + CF[H,E]4 * (1 – CFk2)

= 0.953 + 0 * (1 - 0.953)

= 0.953

So the CF of the symptoms entered by the user for the type of marijuana drug is likely to be 0.953 or 95%.

3. For J-004 = Heroin

G07 = Depression (0.8)

CF[H,E] = CF[H] * CF[E]

= (0.8 * 0.8)

= 0.64

G02 = Anxious and restless (1)

CF[H,E] = CF[H] * CF[E]

= (0 * 1)

= 0

CFk1 = CF[H,E]1 + CF[H,E]2 * (1 - CF[H,E]1)

= 0.64 + 0 * (1 - 0.64)

= 0.64

So the CF of the symptoms entered by the user for the type of heroin drug is most likely 0.64 or 64%.

4. For J-007 = Amphetamine

G01 = Out of breath (1)

CF[H,E]1 = CF[H]1 * CF[E]2

= (0.8 * 1)

= 0.8

So the CF of the symptoms entered by the user for the type of Amphetamine is most likely 0.8 or 80%.

Based on the value of manual calculations, the largest CF value is taken, which is 0.953 or 95% with the type of marijuana drug. It means the patient is using a type of marijuana drug. A case example is tested on an expert system application program (or on machine learning). If the patient's symptoms in the case sample (with the same symptoms) are entered into the expert system built in this study, the result of the process is as shown in Fig. 6.

HASIL KONSULTASI

Nama	:	Andi
Umur	:	24
Jenis Kelamin	:	Laki-laki
Pekerjaan	:	PNS
Alamat	:	ampenan

No	Pertanyaan	Jawaban
1	Sesak Nafas	TIDAK
2	Cemas dan Gelisah	IYA
3	Mual dan Muntah	TIDAK
4	Diare	TIDAK
5	Kejang-kejang	TIDAK
6	Mudah Marah	TIDAK
7	Depresi	IYA

Fig. 6. Screenshot of expert system questions about drug symptoms experienced by patients

Expert system testing on machine learning shows that the expert system has succeeded in correctly identifying the user and the type of drug used by the user. In order to know how accurate the machine learning expertise is, this study also tested several other patients by comparing the results with the urine test results at the Indonesian National Narcotic Agency laboratory in Mataram, Indonesia.

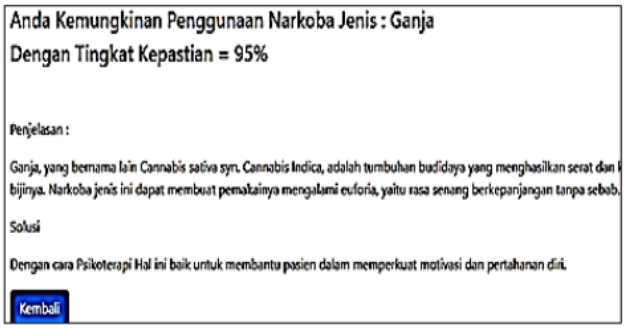

Fig.7. Screenshot of the expert system test results on the type of drug used by the patient

Fig. 7 describes it as follows: You are probably using a type of drug with a 95% certainty. Another narrative in figure 7 is: another name for cannabis is cannabis sativa. Cannabis sativa is a cultivated plant that contains fiber and narcotic substances in its seeds. This drug makes user experience euphoria, namely a prolonged feeling of pleasure for no reason. The cure is psychotherapy, which helps the patient strengthen the motivation to stop using it.

Expert system testing on machine learning shows that the expert system has succeeded in correctly identifying the user and the type of drug used by the user. In order to know how accurate the machine learning expertise is, this study also tested several other patients by comparing the results with the urine test results at the Indonesian National Narcotic Agency laboratory.

C. *Accuracy Test of Machine Learning*

The accuracy test of machine learning in this study is to determine the expert performance of the application system built-in diagnosing users and the types of drugs used. Testing the level of accuracy of machine learning expertise is to compare the suitability of the results with the urine test results from patients at the Indonesian National Narcotics Agency. In a trial of 30 times on 30 patients, there were 30 results of machine learning tests that can correctly identify drug users and as many as 24 machine learning test results that can detect the types of drugs used by drug users. It means that the test results on the data of 30 drug patients show that an expert system on machine learning built using the Certainty Factor method has expertise in diagnosing drug users up to 80 percent. The accuracy rate of up to 80 percent is obtained from the results of the calculation of 24 divided by 30 and multiplied by 100%. Table VI shows the details of the test results in diagnosing the type of abuse of drug users.

TABLE VI. Machine Learning Expertise Accuracy Test Result

Number	Case	System Result	Expert Result	Suitability
1	G01, G02, G03, G04, G05	Cocaine	Cocaine	Suitable
2	G06, G02, G07, G08, G09, G10	Marijuana	Marijuana	Suitable
3	G05, G11, G12, G13, G14	Ecstasy	Ecstasy	Suitable
4	G15, G02, G07, G16	Heroin	Heroin	Suitable
5	G11, G01, G16, G17, G18	Methamphetamine	Methamphetamine	Suitable
6	G09, G11, G18, G19, G10, G14	Hallucinogen	Hallucinogen	Suitable
7	G18, G03, G04, G05, G01, G20	Amphetamines	Amphetamines	Suitable
8	G07, G13, G05, G03	Pethidine	Pethidine	Suitable
9	G27, G03, G18, G21, G22	Codeine	Codeine	Suitable
10	G23, G24, G09, G25, G26	Morphine	Morphine	Suitable
11	G03, G06, G07, G15	Pethidine	Pethidine	Suitable
12	G01, G08, G09, G18	Codeine	Codeine	Suitable
13	G01, G02, G05, G09, G11,G15, G18	Codeine	Hallucinogen	Not suitable
14	G04, G12, G13, G17	Ecstasy	Ecstasy	Suitable
15	G08, G11, G17, G18, G19	Methamphetamine	Methamphetamine	Suitable
16	G01, G13, G15, G20, G22	Amphetamines	Amphetamines	Suitable
17	G17, G20, G21, G22, G23, G25, G27	Codeine	Morphine	Not suitable
18	G02, G07, G11, G15, G16	Pethidine	Heroin	Not suitable
19	G03, G07, G10, G14, G19	Pethidine	Pethidine	Suitable
20	G08, G11, G15, G18, G19, G20, G22, G23	Codeine	Amphetamines	Not suitable
21	G02, G06, G07, G08, G09, G10, G19, G27	Marijuana	Marijuana	Suitable
22	G05, G11, G12, G14, G19, G20	Hallucinogen	Ecstasy	Not suitable
23	G01, G06, G07, G10, G24	Pethidine	Marijuana	Not suitable
24	G03, G07, G09, G21	Morphine	Morphine	Suitable
25	G05, G12, G17, G22	Ecstasy	Ecstasy	Suitable
26	G07, G12, G15, G26	Heroin	Heroin	Suitable
27	G02, G08, G15, G23	Heroin	Heroin	Suitable
28	G06, G09, G15, G26	Morphine	Morphine	Suitable
29	G07, G11, G18, G23, G27	Codeine	Codeine	Suitable
30	G08, G15, G25, G26	Morphine	Morphine	Suitable

IV. CONCLUSION

The results of this study found that: (a). Machine learning in this study can predict drug users and types of drugs based on the symptoms that drug users complain about. (b). This study machine learning acquired knowledge about the symptoms of drug users, types of drugs, and basic knowledge related to the weight of the certainty factor of each type of drug and the symptoms caused so that it can diagnose drug users and the types of drugs used by users. (c). The accuracy of machine learning achieved in this study in predicting the types of drugs used by users and the types of drugs used by users reached 80%. (d). The expert system in this research is website-based, so that the expert system from this research can be used by various parties and in different places to identify users and the types of drugs used by users.

The implication of this research result is that the expert system built in this study can be a tool (choice) to replace or complete the testing system for drug users through urine testing in the laboratory.

The drawback of the results of this study is that machine learning expertise in this study is only limited to simple machine learning, as is the case with simple machine learning which was built on previous research by Zhongheng Zhang (2016), which used the KNN method in building learning machines. Furthermore, the machine learning expertise generated from this research is only limited to the expertise possessed in accordance with the knowledge obtained (symptoms, types of drug abuse, rule base, and calculation of certainty factor) under study. Therefore, further research needs to build machine learning that can increase its expertise based on more new data and use another method.

REFERENCES

[1] R. Jiménez, J. Anupol, B. Cajal, and E. Gervilla, "Data mining techniques drug use research," *Addict. Behav. Reports*, vol. 8, pp. 128–135, 2018.

[2] F. M. Vassoler, E. M. Byrnes, and R. C. Pierce, "The impact of exposure to addictive drugs on future generations: Physiological and behavioral effects," *Neuropharmacology*, vol. 76, no. PART B, pp. 269–275, 2014.

[3] P. K. Shanmugam, "The Influence of Social Factors in Drug Addiction–A Mini Review of Work by Miller & Carroll (2006)," *J. Alcohol. Drug end.*, vol. 05, no. 04, pp. 4–6, 2017.

[4] Boys, J. Marsden, and J. Strang, "Understanding reasons for drug use amongst young people: A functional perspective," *Health Educ. Res.*, vol. 16, no. 4, pp. 457–469, 2001.

[5] G. López, L. M. Orchowski, M. K. Reddy, J. Nargiso, and J. E. Johnson, "A review of research-supported group treatments for drug use rders," *BMC Public Health*, vol. 16, no. 51, pp. 1–21, 2021.

[6] Z. Justinova, L. V. Panlilio, and S. R. Goldberg, "Drug Addiction," *Natl. Libr. Medicine*, vol. 1, no. 1, pp. 310–335, 2009.

[7] T. Saah, "The evolutionary origins and significance of drug addiction," *Harm Reduct. J.*, vol. 2, pp. 1–7, 2005.

[8] G. Leshner M.Stevens, S. Kim, N. Kim, T. L.Wagener, and A. C. Villantie, "Cognitive and affective responses to marijuana prevention and educational messaging," *Drug Alcohol Depend.*, vol. 225, no. August, pp. 1–3, 2021.

[9] M. Zaman *et al.*, "Drug abuse among the students," *Pakistan J. Pharm.* , vol. 1, no. 1, p. 41, 2015.

[10] L. W. Chou, K. M. Chang, and I. Puspitasari, "Drug Abuse Research

Trend Investigation with Text Mining," *Comput. Math. Methods Med.*, vol. 2020, pp. 1–8, 2020.

[11] A. Bevan and N. Patel, "An Electronic Prescription Alerting System-Improving the Discharge Medicines Process," *Arch. Dis. Child.*, vol. 101, no. 9, p. e2.55-e2, 2016.

[12] A. Tsyben, N. Gooding, and W. Kelsall, "Assessing the Impact of a Newly Introduced Electronic Prescribing System Across a Paediatric Department – Lessons Learned," *Arch. Dis. Child.*, vol. 101, no. 9, p. 3-e2, 2016.

[13] M. Mahmoudi, S. Pakpour, and G. Perry, "Drug-Abuse Nanotechnology: Opportunities and Challenges," *ACS Chem. Neurosci.*, vol. 9, no. 10, pp. 2288–2298, 2018.

[14] J. Redman, "Recognizing the Warning Signs of Drug Addiction : What You Need to Know," *Mental Health and Counseling Studies*. pp. 1–7, 2021.

[15] N. Jojen, "A Survey Paper on Data Mining Techniques in Drug Industry," *. J. Eng. Res. Technol.*, vol. 3, no. 30, pp. 296–299, 2015.

[16] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or n category," *Front. Chem.*, vol. 6, no. May, pp. 1–11, 2018.

[17] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. nsl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.

[18] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 4, pp. 319–326, 2021.

[19] A. Anggrawan, K. Hidjah, and Q. S. Jihadil, "Kidney failure diagnosis based on case-based reasoning (CBR) method and statistical analysis," in *2016 International Conference on Informatics and Computing, ICIC 16*, 2017, pp. 298–303.

[20] K. Muludi, R. Suharjo, A. Syarif, and F. Ramadhani, "Implementation of forward chaining and certainty factor m d on android-based expert stem of tomato diseases identification," *Int. J. Adv. Comput. Sci. Appl.*, l. 9, no. 9, pp. 451–456, 2018.

[21] C. P. C. Munaiseche, D. R. Kaparang, and P. T. D. Rompas, "An Expert System for Diagnosing Eye Diseases using Forward Chaining Method," in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. no. 1, pp. 1–8.

[22] N. Von Greiff and L. Skogens, "Recovery and identity: a five-year follow-up of persons treated in 12-step-related programs," *Drugs Educ. ev. Policy*, pp. 1–10, 2021.

[23] A. Anggrawan, N. Ibrahim, S. Muslim, and C. Satria, "Interaction between Learning Style and Gender in Mixed Learning with 40 % Face-to-face Learning and 60 % Online Learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, pp. 407–413, 2019.

Similarity Check _ Machine Learning for Diagnosing Drug Users and Types of Drugs used

ORIGINALITY REPORT

15%
SIMILARITY INDEX

PRIMARY SOURCES

1	repository.universitasbumigora.ac.id Internet	244 words — 5%
2	www.jatit.org Internet	55 words — 1%
3	Mohammed Jasim A. Alkhafaji, Abbas Fadhil Aljuboori, Abdullahi Abdu Ibrahim. "Clean medical data and predict heart disease", 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2020 Crossref	47 words — 1%
4	journal.uad.ac.id Internet	39 words — 1%
5	arxiv.org Internet	36 words — 1%
6	onlinelibrary.wiley.com Internet	32 words — 1%
7	www.ijrte.org Internet	30 words — 1%
8	Joan Angelina Widians, Novianti Puspitasari, Aidil Febriansyah. "Disease Diagnosis System Using	26 words — < 1%

Certainty Factor", 2019 International Conference on Electrical, Electronics and Information Engineering (ICEEIE), 2019
Crossref

9	turcomat.org Internet	26 words — < 1%
10	Arifin Ika Nugroho, Pulung Riyanto, Jori Lahinda, Serli Hatul Hidayat. "Effectiveness of physical condition training on the fitness of drug abuse victims, BRSPP Yogyakarta", Enfermería Clínica, 2020 Crossref	20 words — < 1%
11	www.ijeat.org Internet	20 words — < 1%
12	academic.oup.com Internet	19 words — < 1%
13	www.tandfonline.com Internet	18 words — < 1%
14	Anthony Anggrawan, Nurdin Ibrahim, Suyitno Muslim, Christofer Satria. "Interaction between Learning Style and Gender in Mixed Learning with 40% Face-to-face Learning and 60% Online Learning", International Journal of Advanced Computer Science and Applications, 2019 Crossref	17 words — < 1%
15	publikasi.mercubuana.ac.id Internet	16 words — < 1%
16	Ashok Singh. "Novel nanoparticle-based treatment approaches", Elsevier BV, 2021 Crossref	15 words — < 1%
17	Edgar A. Whitley. "Two approaches to developing expert systems: A consideration of formal and	15 words — < 1%

semi-formal domains", AI & Society, 1991
Crossref

18 Ania Wajnberg, Katherine Ornstein, Meng Zhang, Kristofer L. Smith, Theresa Soriano. "Symptom Burden in Chronically Ill Homebound Individuals", Journal of the American Geriatrics Society, 2013 — 14 words — < 1%
Crossref

19 W Uriawan, K Manaf, U Syaripudin, M Mujiburrahman. "Expert system for diagnosing narcissistic personality disorders using certainty factor and forward chaining methods", IOP Conference Series: Materials Science and Engineering, 2021 — 14 words — < 1%
Crossref

20 Win Min Thit, Sai Wai Yan Myint Thu, Jaranit Kaewkungwal, Ngamphol Soonthornworasiri et al. "User Acceptance of Electronic Medical Record System: Implementation at Marie Stopes International, Myanmar", Healthcare Informatics Research, 2020 — 13 words — < 1%
Crossref

21 repositorio.unal.edu.co — 12 words — < 1%
Internet

22 Wu, L.T.. "Hallucinogen-related disorders in a national sample of adolescents: The influence of ecstasy/MDMA use", Drug and Alcohol Dependence, 20090901 — 11 words — < 1%
Crossref

23 Hengki Tamando Sihotang, Fristi Riandari, R.Mahdalena Simanjorang, Agustina Simangunsong, Penda Sudarto Hasugian. "Expert System for Diagnosis Chicken Disease using Bayes Theorem", Journal of Physics: Conference Series, 2019 — 10 words — < 1%
Crossref

24	Zhang, Rui-san, Zhen He, Wei-dong Jin, and Rui Wang. "Effects of the cannabinoid 1 receptor peptide ligands hemopressin, (m)RVD-hemopressin(?) and (m)VD-hemopressin(?) on memory in novel object and object location recognition tasks in normal young and A?1-42-treated mice", Neurobiology of Learning and Memory, 2016. Crossref	10 words —	< 1%
25	Li Ju, Aihua Lyu, Hongxia Hao, Wen Shen, Hua Cui. "Deep Learning-Assisted Three-Dimensional Fluorescence Difference Spectroscopy for Identification and Semiquantification of Illicit Drugs in Biofluids", Analytical Chemistry, 2019 Crossref	9 words —	< 1%
26	Glenn Leshner, Elise M. Stevens, Amy M. Cohn, Seunghyun Kim, Narae Kim, Theodore L. Wagener, Andrea C. Villanti. "Cognitive and affective responses to marijuana prevention and educational messaging", Drug and Alcohol Dependence, 2021 Crossref	8 words —	< 1%
27	Sutrisno Arianto Pasaribu, Poltak Sihombing, Suherman Suherman. "Expert System for Diagnosing Dental and Mouth Diseases with a Website-Based Certainty Factor (CF) Method", 2020 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT), 2020 Crossref	8 words —	< 1%
28	ojs.trigunadharma.ac.id Internet	7 words —	< 1%
29	Deny Adhar, Syafrizal, Erwin Panggabean, Lahmudin Sipahutar, Muhammad Reza Fahlevi. "Expert System to Diagnose Gastritis Disease Using the	6 words —	< 1%

Certainty Factor Method", 2019 7th International Conference on Cyber and IT Service Management (CITSM), 2019
Crossref

30 Ninive von Greiff, Lisa Skogens. "Recovery and identity: a five-year follow-up of persons treated in 12-step-related programs", Drugs: Education, Prevention and Policy, 2021 — 6 words — < 1%
Crossref

EXCLUDE QUOTES	OFF	EXCLUDE MATCHES	OFF
EXCLUDE BIBLIOGRAPHY	OFF		

IJACSA Acceptance Notification - Volur

Editor IJACSA
to me, chris, cknuraini, lusiana, ayu.areyu, mayadi.yadot

Dear Author(s),

Congratulations, your submitted paper titled "Machine Learning for D
Journal of Advanced Computer Science and Applications (IJACSA).

Registration and Publication Fee Payment
You may now proceed with the registration for paper publication at h
please get in touch with us.

Kindly register before **19 November 2021** for timely publication and

Open Access
All papers published by the International Journal of Advanced Comp
any subscription charges. All published articles are assigned a DOI.

Indexing
Upon publication of papers, our next step will be to submit all publisl
(Q3), Inspec, Ebesco, ProQuest, Microsoft Academic, WorldCat. IJA

Reviewer Feedback
All submitted manuscripts were subject to a double-blind peer reviev

Mail

Inbox 1,392

Starred

Snoozed

Chat

No conversations
Start a chat

Spaces

No spaces yet
Create or find a space

Meet

University/ Organization	Universitas Bumigora
Country	Indonesia
Paper Title	Machine Learning for Diag
Registration Amount	USD 450

Payment Details

Payment ID	pay_ILNoF

Please Note: The charges will be billed to your credit or debit card a

Thank you for your mail. Thanks a lot. Noted

Plagiarism check _ camera ready _ Machine Learning for Diagnosing Drug Users and Types of Drugs used

By Anthony Anggrawan

WORD COUNT	5387	TIME SUBMITTED	17-NOV-2021 01:08PM
		PAPER ID	78938767

Machine Learning for Diagnosing Drug Users and Types of Drugs used

Anthony Anggrawan[1]
Ni Gusti Ayu Dasriani[5]
Mayadi[6]
Computer Science Study Program
Bumigora University, Mataram, Indonesia

Christofer Satria[2]
Visual Communication Design Study Program
Bumigora University, Mataram, Indonesia

Che Ku Nuraini[3]
Department of Centre of Research and Innovation Management
Universiti Teknikal Malaysia Melaka, Malaysia

Lusiana[4]
Technical Information Study Program
STMIK AMIK Riau, Riau, Indonesia

***Abstract*— Drug use is very detrimental to the physical and psychological health of users. Drug abuse also causes addiction and is a global epidemic. Therefore it is not surprising that scientific research related to drugs has attracted attention for research. However, many factors become obstacles in the medical services of the drug user, including cost, flexibility, and a slow process. Meanwhile, electronic systems can speed up handling time, improve work efficiency, save costs and reduce inspection errors. It means that a breakthrough is needed in developing a platform that can identify drug users. Therefore, this research aims to build machine learning with expertise like an expert who can diagnose drug users and distinguish the types of drugs used by drug users. The expert system on machine learning was developed using the Forward Chaining and Certainty Factor methods. This study concludes that the expert system on machine learning developed can be used to diagnose drug users and distinguish the types of drugs used with an accuracy of up to 80%. The implications of the expert system on machine learning are an alternative method for narcotics officers and medical doctors in diagnosing drug users and the types of drugs used.**

***Keywords*—Machine learning; Drug; Expert system; Forward chaining; Certainty factor**

I. INTRODUCTION

Social environment factors have influenced others to engage in drug use [1] [2]. Adult figures who are addicted to drugs have a great influence on the behavior of others to become addicted [3]. Poor, low skills, life pressure, anxiety, and deviant behavior are factors that also lead to drug use [3]. Relaxing, drinking, staying up late, increasing enthusiasm, and relieving stress are other triggering factors for many drug use among young people [4]. Drug use, including amphetamines, marijuana, cocaine, heroin, and the like, are a major public health problem in physiological symptoms, resulting in behavioral changes, cognitive problems, and mental health [5]. Drug use also affects the physiology and behavior of future generations [2].

Drug abuse causes physical dependence (addiction) or relapse to continuously consume [6], although it has resulted in physical and psychological problems [6]. The previous research shows that drug users are very high [7] [1] and increasing globally [8]. Drug abuse has become a global epidemic that affects human behavior [9]. Because drug use is very detrimental to the physical and psychological harm of the user, it is not surprising that this research related to drugs has attracted attention for scientific research [10].

Another factor that is often considered in medical services is the cost factor and its inflexibility (Bevan & Patel, 2016). Processes that are done manually tend to cause delays in medical diagnosis [11]. Using an electronic system can speed up handling time, improve work efficiency, and save costs [11]. Using an electronic system allows lower errors and eliminates omissions in deciding the test results and achieving the results [12]. However, the success of curing drug abuse and dependence is still limited; this includes the lack of success in the early identification of at-risk populations, resulting in increased death rates due to overdose [13]. In other words, not paying attention to the early symptoms of consuming drugs will be disastrous and make people who are loved suffer the destructive effects of the substance [14].

Meanwhile, if it turns out to be able to identify it early, it can prevent harmful consequences in the future that are sure to occur [14]. It means that there is a need for breakthroughs in developing platforms that can identify and screen patients susceptible to addiction after using opioid drugs [13]. Therefore, this research aims to develop a machine learning that has expertise like an expert. The expert system created can identify and screen or diagnose early drug users and the types of drugs used by using the Certainty Factor and Forward Chaining methods. The certainty factor method measures the certainty of the type of drug used by the user or patient who conducts consultations. On the other hand, forward chaining plays a role in the flow of the reasoning process from beginning to end based on data mining of physical symptoms of drug users and the types of drugs used (collected or explored previously).

Medical data is helpful as the knowledge that helps make scientific decisions regarding drug use [15]. Electronic medical use based on doctor's notes is useful for an effective treatment medium [10]. In the meantime, data mining is capable of electronic checks based on the patient's medical record [10].

Besides, the machine learning methods are a technique that can be useful for finding correlations based on the case for prediction purposes [16]. Unfortunately, machine learning is still few in the medical field due to technical problems [17]. Therefore, the simple machine learning method built in this research by imitating (studying) human knowledge in analyzing the physical symptoms of drug users and then implementing it in predicting drug users and identifying drug types used by users. It means that the expert system in machine learning has intelligence like an expert in diagnosing users and the types of drugs used from the physical symptoms that arise from drug users. Taking into account that the current use of information and communication technology (ICT) is growing or expanding very quickly or booming [18]. Therefore, the embodiment of the machine learning system in this research is website-based. So, anyone (the public) can access it from anywhere and has flexibility because it can work on various devices and operating systems. Therefore, a machine learning system in this research is helpful for early diagnosis without having to examine a narcotics laboratory and without a doctor or expert.

It is necessary to know the percentage of machine learning efficacy in identifying drug users and the type of drug used. It means that further testing to determine the actual percentage of machine learning efficacy still needs to be done. This study makes this happen by comparing the test results achieved by machine learning based on symptoms of drug users compared to the test results achieved from laboratory tests of drug users' urine in identifying drug users and the type of drug used. If machine learning has high efficacy, it can save time and cost of drug testing for suspects or drug users by using machine learning compared to testing drugs on urine or blood for suspects and drug users.

Some recent works related to this research:

- Zhongheng Zhang (2016) introduced the k-nearest neighbor (kNN) method as a simple machine learning method for modeling [17]. The similarity between the research in this article and the previous one is that they both use a simple approach to machine learning. While the difference is that the research uses the certainty factor method and forward chaining for machine learning, while previous research uses the kNN method for machine learning. Another difference is that the previous research was focus on predicting the class from the new dataset to the most similar class. In contrast, the research in this article focuses on machine learning to diagnose drug users and the types of drugs used.
- Anthony Anggrawan, Khasnur Hidjah, and Jihadil Qudsi S. (2017) implement intelligent application programs to detect kidney failure [19]. The previous research and the research in this article have similarities in developing web application programs with the PHP programming language and MySQL database. In addition, the last analysis used medical data on failure cases to diagnose new renal illness issues using CBR (Case-Based Reasoning method). In contrast, the articles in this study use the expertise of drug experts (specialists) as knowledge of the application system for early diagnosis of drug users and the types of drugs used by drug users using the Forward Chaining and Certainty Factor methods.
- Kurnia Muludi, Radix Suharjo, Admi Syarif, and Fitria Ramadhani (2018) identified tomato plant diseases [20]. This previous research and the research in this article both implements forward chaining and certainty factor methods. However, the difference in the last research is to build an expert system to identify plant diseases based on android [20]. In contrast, the research in the article builds an expert system to identify users of drugs and the types of drugs used based on the website.
- Munaiseche, Kaparang, and Rompas (2018) built an expert system to assist doctors in diagnosing eye diseases [21]. In contrast to the research in this article, it is to create an expert system to diagnose drug users and the types of drugs used. Furthermore, this previous research used the forward chaining method, while the research in this article uses the forward chaining method and certainty factor. The similarities between the previous study and the research in this article are that both use PHP and MySQL in building an expert system.
- Ninive Von Greiff and Lisa Skogens (2021) investigated a drug user recovery program for drug addiction [22]. The research method is the interview or qualitative approach [22]. The similarity of the research in this article with previous studies is that they both study drug users. The difference is in previous studies examining the results of addiction recovery on drugs with the interview method. Meanwhile, the research in this article builds machine learning that has an intelligent system to detect drug users and the types of drugs used.

The latest related work identifies that the article in this study has a novelty that no previous researcher has researched. Another strength of this research is conducting a comparative test to determine the efficacy of machine learning or expert systems developed in identifying users and the types of drugs used by users that have not been studied before.

The systematics of writing this paper is as follows: the following sub-section discusses the research methodology, which includes research data and research methods used. The next subsection discusses the results and discussion of the research. Finally, the conclusions obtained from the study results and suggestions for further research are narrated in the Conclusions subsection.

II. Research Methodology

This study is a case study at the Indonesian National Narcotics Agency (*Badan Narkotika Nasional* or BNN) in Mataram, Indonesia. The number of drug users used as samples to test the expertise and accuracy of the machine learning built in this study was 30. The selection of data samples in this study was random. This research's development of machine learning expertise consists of stages: knowledge acquisition, expert system design (programming), machine learning/expert system testing, and accuracy test (see Fig.1).

A. Knowledge Acquisition

For the system development stage, the effort made is to obtain knowledge from drug experts, which is used as a knowledge base to build the machine learning or expert systems. The method used in obtaining knowledge related to narcotics is the interview method. The knowledge gained is the

knowledge about the types of drugs and their symptoms. Based on the knowledge obtained, there are ten types of drugs that drug users dominantly use, and there are 27 types of drug symptoms

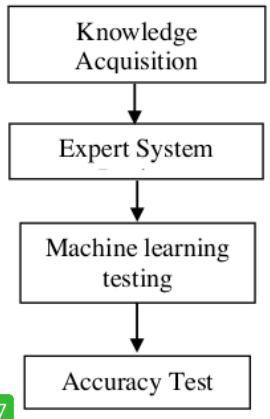

Fig.1. Stages in the development of an expert system on machine learning

In the knowledge acquisition stage, an expert from the Indonesian National Narcotics Agency (in Mataram, Indonesia) shares knowledge about drug use, including those related to symptoms and the types of drugs used by drug users. The knowledge gained at this stage serves as a knowledge base in building expertise from machine learning.

B. *Expert System Design*

The expert system design stage is a process for modeling the data that has been collected and designing an application system that is planned according to programming problems and the acquisition of knowledge obtained. This stage in computer science is known as planning the use case diagram design, data flow diagram (DFD) design, database design, and flowchart to be built on the application program. The programming stage is the implementation stage of the system design plan into a computer programming language [23]. This research uses the PHP and MySQL programming languages. The computer application program that is built is a cloud application program. Application development with PHP programming language and MYSQL database makes application programs can run via the web. By being stored on the server computer, the application program is ubiquitous. The ubiquitous application program means that the application program can be accessed from anywhere and at any time [24].

C. *Machine learning testing*

The machine learning testing phase is the functional testing phase of the built application or black-box testing. Black-box testing is a test that no longer involves programming code or programming languages. In short, black-box testing on the expert system in this study is to determine whether the expert system built is under the list of desired system requirements.

D. *Accuracy Test*

Testing accuracy in machine learning is to determine the level of expertise of the expert system built in this study. It means that it is known how much accuracy the expert system has built-in diagnosing users' types of narcotics.

III. Result and Discussion

A. *Knowledge Acquisition*

The knowledge acquisition stage is the stage of acquiring the required knowledge data. The acquired knowledge acquisition data is useful in solving programming logic in diagnosing users and the types of drugs used by drug users. Table I shows the code for the type of drug used by drug users. Meanwhile, Table II presents the code of symptoms caused by drug users.

TABLE I. List of Types of Drugs

No.	Drug Type Code	Drug Name
1	P01	Cocaine
2	P02	Marijuana
3	P03	Ecstasy
4	P04	Heroin
5	P05	Methamphetamine
6	P06	Hallucinogen
7	P07	Amphetamine
8	P08	Pethidine
9	P09	Codeine
10	P10	Morphine

TABLE II. List of Symptom of Drug User

Symptoms of Drug User			
Code	**Symptom**	**Code**	**Symptom**
G01	Out of breath	G15	Difficult to focus
G02	Anxious and restless	G16	Difficult to rest
G03	Nausea and vomiting	G17	Weight loss
G04	Diarrhea	G18	Dry mouth
G05	Convulsions	G19	Blurred vision
G06	Easy to get angry	G20	Changes in skin color
G07	Depression	G21	Constipation
G08	Changes in sleep patterns	G22	Stomachache
G09	Sweating	G23	Drowsiness
G10	Chills (Hot cold)	G24	Itching
G11	Shaking	G25	Difficulty urinating
G12	Insomnia	G26	Mood swings
G13	Fast heart rate	G27	Dizziness
G14	Increased blood pressure		

TABLE III. Rule Base of Types and Symptoms of Basic Diagnosis

Rule Base	
Drug Type	**Symptom**
P01	G01 and G02 and G03 and G04 and G05
P02	G06 and G02 and G07 and G08 and G09 and G10
P03	G05 and G11 and G12 and G13 and G14
P04	G15 and G02 and G07 and G16
P05	G11 and G01 and G16 and G17 and G18
P06	G09 and G11 and G18 and G19 and G10 and G14
P07	G18 and G03 and G04 and G05 and G01 and G20
P08	G07 and G13 and G05 and G03
P09	G27 and G03 and G18 and G21 and G22
P10	G023 and G24 and G09 and G25 and G26

After modeling the acquired knowledge acquisition data or knowledge representation (as shown in Table III) the next step is to implement it into the certainty factor algorithm. The certainty factor uses a value between 0.2 and 1.0 to assume a level of confidence in the data. A simulation of the calculation of the certainty factor was carried out based on the weight of symptoms arising from the type of drug used by drug users with weights of 0.8 and 1.0 according to the opinion of the drug expert (See Table IV).

TABLE IV. DETERMINATION OF DRUG SYMPTOM WEIGHT SCORE ACCORDING TO THE DECISION OF DRUG EXPERTS

No	Symptom	Weight Score
1	Very often	1
2	Often	0,8
3	Never	0

So, on the drug symptom weighted score given to the certainty factor, a score of 0 indicates that drug users do not experience these symptoms. If a drug user experiences symptoms, then the weighted score given for the frequently experienced symptoms is 0.8 and the most frequently experienced is 1.0, according to the drug expert's decision.

TABLE V. KNOWLEDGE BASE

Symptom	J-001	J-002	J-003	J-004	J-005	J-006	J-007	J-008	J-009	J-010
	CF	CF	CF	CF	CF	CF	CF	CF	CF	CF
Out of breath					0,8		1			
Anxious and restless	0,8	0,8		1						
Nausea and vomiting	0,8						0,8	0,8	0,8	
Diarrhea	1						0,8			
Convulsions	0,8		0,8				0,8	0,8		
Easy to get angry		1								
Depression		0,8		0,8				1		
Changes in sleep pattams		0,8		0,8						
Sweating		0,8				0,8				0,8
Chills		0,8				0,8				
Shaking			0,8		0,8	0,8				
Insomnia			0,8							
Fast heart rate			0,8					0,8		
Increased blood pressure			0,8			1				
Difficult to focus				1						
Difficult to rest					0,8					
Weight loss					0,8					
Dry mouth					1	0,8	0,8		1	
Blurred vision						0,8				
Changes in skin color							0,8			
Constipation									0,8	
Stomachache									0,8	
Drowsiness										1
Itching										0,8
Difficulty urinating										0,8
Mood swings										1
Dizziness									0,8	

This study's knowledge base of machine learning expertise is the symptoms, types of drugs, and CF rules obtained from drug experts (see Table V). The knowledge base is an essential component that contains the knowledge possessed by competent experts in the related field (i.e., narcotics in this study). Furthermore, the knowledge base is the basis for decision-making in an expert system, where this decision-making is related to the process of retrieving previously collected and stored knowledge.

B. Expert System Design

Figure 2 shows the interactions that occur between system users and the developed expert system.

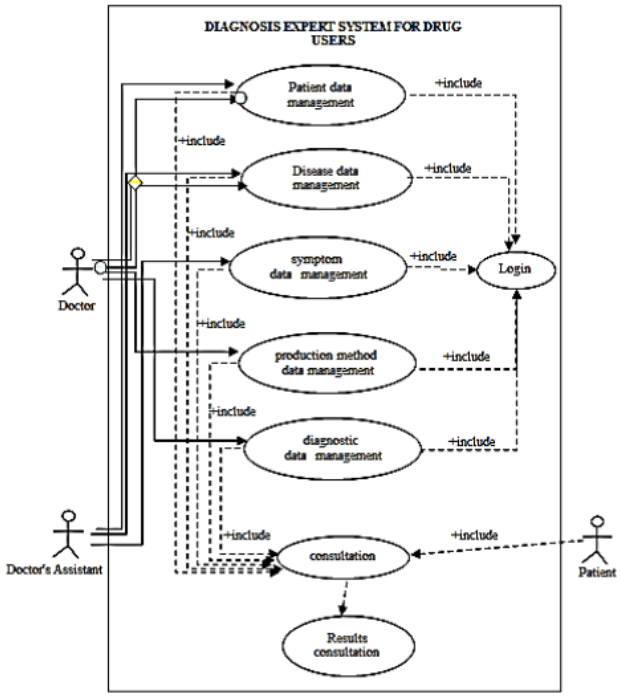

Fig. 2. Use Case Diagram on machine learning

This study has a database that stores records of users, patients, symptoms, and types of drugs, including diagnostic data, so it is necessary to design a data workflow model to realize a structured program.

The Data Flow Diagram (DFD) in Fig. 3 and Fig. 4 illustrates where the data flow comes from and where the data processing on the expert system is built.

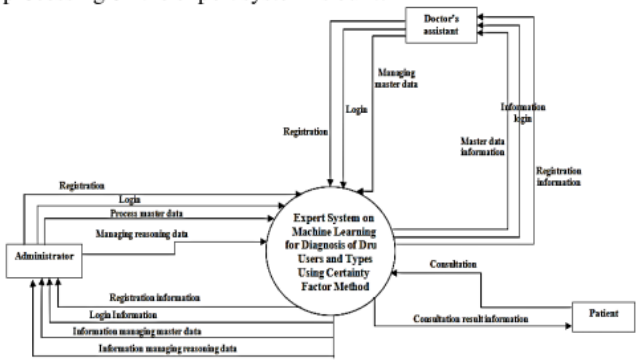

Fig. 3. Context Diagram of Data Flow on Machine Learning

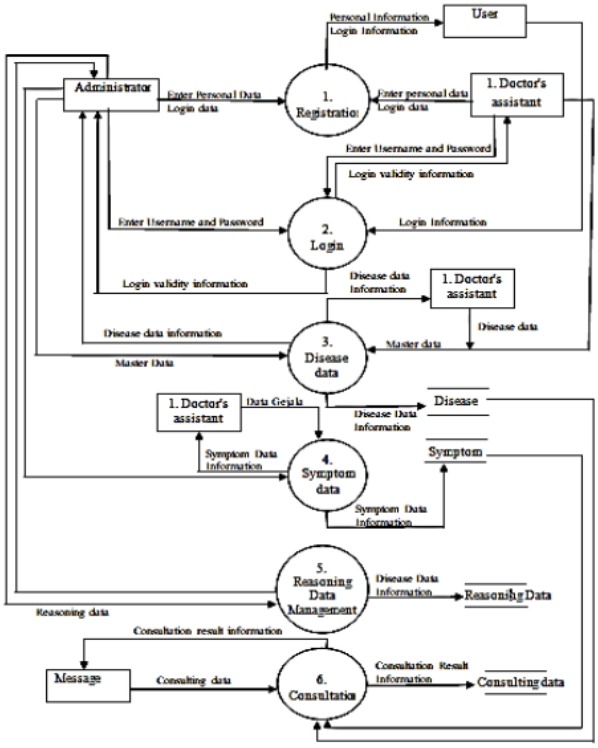

Fig. 4. Overview Diagram of Data Flow on Machine Learning

The context diagram in Fig. 3 shows the data flow of the system globally. In contrast, the overview diagram in Figure 4 shows a more detailed data flow that the system performs and its engagement with external data.

The flow diagram in fig. 5 shows a series of flow relationships in the expert system built in this study or shows the overall process sequence in building an expert system in this study.

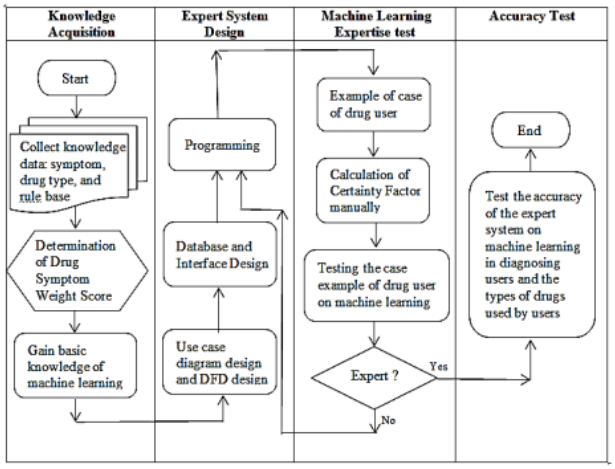

Fig. 5. Flow diagram of the whole Process of Building an Expert System on the Machine Learning

The flow diagram contains a more detailed description of how each step of the procedure is actually carried out in building an expert system on machine learning that can diagnose users and the types of drugs used by users.

C. Machine learning testing

Expertise testing of machine learning is carried out using case samples from a patient. For example, in one case, a drug patient had symptoms of shortness of breath, depression, chills, anxiety and restlessness, and irritability. Symptoms of drug patients who have symptoms of shortness of breath, depression, chills, anxiety and restlessness, and irritability are symptoms of drug users: (1) Cocaine, (2) Cannabis, (3) Heroin, and (4) Amphetamine.

The formula for CF is:

$CF[H,E] = CF[H] * CF[E]$

$CF\ Combine\ CF[H,E]1 = CF[H,E]1 + CF[H,E]\ 2 * (1 - CF[H,E]\ 1)$

$CF\ Combine\ CF[H,E]\ old3 = CF[H,E]\ old + CF[H,\ E]\ * (1 - CF[H,E]\ old)$

Based on manual calculations, the results are as follows:

1. For J-001 = Cocaine

   G01 = Out of breath (0.8)

   $CF[H,E] = CF[H] * CF[E]$

   $= (0.8 * 0.8)$

   $= 0.64$

   G02 = Anxious and restless (0.8)

   $CF[H,E] = CF[H] * CF[E]$

   $= (0.8 * 0.8)$

   $= 0.64$

   $CFk1 = CF[H,E]\ 1 + CF[H,E]2 * (1 - CF[H,E]1)$

   $= 0.64 + 0.64 * (1 - 0.64)$

   $= 0.870$

   So the expert CF from the symptoms entered by the user for the type of drug Cocaine is probably 0.870 or 87%.

2. For J-002 = Marijuana

   G07 = Depression (0,8)

   $CF[H,E] = CF[H] * CF[E]$

   $= (0.8 * 0.8)$

   $= 0.64$

   G10 = Chills (0.8)

   $CF[H,E] = CF[H] * CF[E]$

   $= (0.8 * 0.8)$

   $= 0.64$

   G02 = Anxious and restless (0.8)

   $CF[H,E] = CF[H] * CF[E]$

   $= (0.8 * 0.8)$

   $= 0.64$

   G06 = Easy to get angry (1)

   $CF[H,E] = CF[H] * CF[E]$

   $= (0 * 1)$

   $= 0$

   $CFk1 = CF[H,E]1 + CF[H,E]2 * (1 - CF[H,E]1)$

= 0.64 + 0.64 * (1 - 0.64)

= 0.870

CFk2 = CFk1 + CF[H,E]3 * (1 - CFk1)

= 0.870 + 0.64 * (1 - 0.870)

= 0.953

CFk3 = CFk2 + CF[H,E]4 * (1 – CFk2)

= 0.953 + 0 * (1 - 0.953)

= 0.953

So the CF of the symptoms entered by the user for the type of marijuana drug is likely to be 0.953 or 95%.

3. For J-004 = Heroin
G07 = Depression (0.8)

CF[H,E] = CF[H] * CF[E]

= (0.8 * 0.8)

= 0.64

G02 = Anxious and restless (1)

CF[H,E] = CF[H] * CF[E]

= (0 * 1)

= 0

CFk1 = CF[H,E]1 + CF[H,E]2 * (1 - CF[H,E]1)

= 0.64 + 0 * (1 - 0.64)

= 0.64

So the CF of the symptoms entered by the user for the type of heroin drug is most likely 0.64 or 64%.

4. For J-007 = Amphetamine
G01 = Out of breath (1)

CF[H,E]1 = CF[H]1 * CF[E]2

= (0.8 * 1) = 0.8

So the CF of the symptoms entered by the user for the type of Amphetamine is most likely 0.8 or 80%.

Based on the value of manual calculations, the largest CF value is taken, which is 0.953 or 95% with the type of marijuana drug. It means the patient is using a type of marijuana drug. A case example is tested on an expert system application program (or on machine learning). If the patient's symptoms in the case sample (with the same symptoms) are entered into the expert system built in this study, the result of the process is as shown in Fig. 6.

HASIL KONSULTASI

Nama	:	Andi
Umur	:	24
Jenis Kelamin	:	Laki-laki
Pekerjaan	:	PNS
Alamat	:	ampenan

No	Pertanyaan	Jawaban
1	Sesak Nafas	TIDAK
2	Cemas dan Gelisah	IYA
3	Mual dan Muntah	TIDAK
4	Diare	TIDAK
5	Kejang-kejang	TIDAK
6	Mudah Marah	TIDAK
7	Depresi	IYA

Fig. 6. Screenshot of expert system questions about drug symptoms experienced by patients

Expert system testing on machine learning shows that the expert system has succeeded in correctly identifying the user and the type of drug used by the user. In order to know how accurate the machine learning expertise is, this study also tested several other patients by comparing the results with the urine test results at the Indonesian National Narcotic Agency laboratory in Mataram, Indonesia.

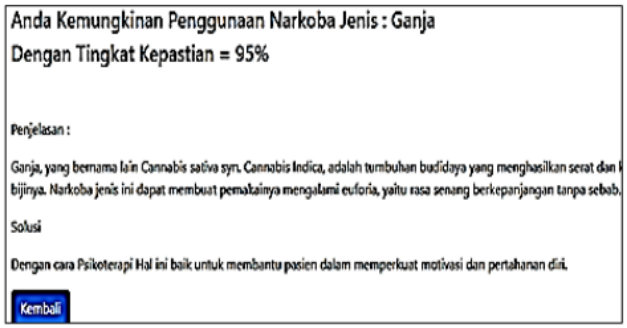

Fig.7. Screenshot of the expert system test results on the type of drug used by the patient

Fig. 7 describes it as follows: You are probably using a type of drug with a 95% certainty. Another narrative in figure 7 is: another name for cannabis is cannabis sativa. Cannabis sativa is a cultivated plant that contains fiber and narcotic substances in its seeds. This drug makes user experience euphoria, namely a prolonged feeling of pleasure for no reason. The cure is psychotherapy, which helps the patient strengthen the motivation to stop using it.

Expert system testing on machine learning shows that the expert system has succeeded in correctly identifying the user and the type of drug used by the user. In order to know how accurate the machine learning expertise is, this study also tested several other patients by comparing the results with the urine test results at the Indonesian National Narcotic Agency laboratory.

D. Accuracy Test of Machine Learning

The accuracy test of machine learning in this study is to determine the expert performance of the application system built-in diagnosing users and the types of drugs used.

Testing the level of accuracy of machine learning expertise is to compare the suitability of the results with the urine test results from patients at the Indonesian National Narcotics Agency. In a trial of 30 times on 30 patients, there were 30 results of machine learning tests that can correctly identify drug users and as many as 24 machine learning test results that can detect the types of drugs used by drug users. It means that the test results on the data of 30 drug patients show that an expert system on machine learning built using the Certainty Factor method has expertise in diagnosing drug users up to 80 percent. The accuracy rate of up to 80 percent is obtained from the calculation results of 24 divided by 30 and multiplied by 100%. The details of the machine learning accuracy test results are as described in Table VI. Table VI shows the comparison between the results of the expert system and the results of drug experts on the diagnosis of the types of drugs used by drug users.

TABLE VI. Machine Learning Expertise Accuracy Test Result

Number	Case	System Result	Expert Result	Suitability
1	G01, G02, G03, G04, G05	Cocaine	Cocaine	Suitable
2	G06, G02, G07, G08, G09, G10	Marijuana	Marijuana	Suitable
3	G05, G11, G12, G13, G14	Ecstasy	Ecstasy	Suitable
4	G15, G02, G07, G16	Heroin	Heroin	Suitable
5	G11, G01, G16, G17, G18	Methamphetamine	Methamphetamine	Suitable
6	G09, G11, G18, G19, G10, G14	Hallucinogen	Hallucinogen	Suitable
7	G18, G03, G04, G05, G01, G20	Amphetamines	Amphetamines	Suitable
8	G07, G13, G05, G03	Pethidine	Pethidine	Suitable
9	G27, G03, G18, G21, G22	Codeine	Codeine	Suitable
10	G23, G24, G09, G25, G26	Morphine	Morphine	Suitable
11	G03, G06, G07, G15	Pethidine	Pethidine	Suitable
12	G01, G08, G09, G18	Codeine	Codeine	Suitable
13	G01, G02, G05, G09, G11,G15, G18	Codeine	Hallucinogen	Not suitable
14	G04, G12, G13, G17	Ecstasy	Ecstasy	Suitable
15	G08, G11, G17, G18, G19	Methamphetamine	Methamphetamine	Suitable
16	G01, G13, G15, G20, G22	Amphetamines	Amphetamines	Suitable
17	G17, G20, G21, G22, G23, G25, G27	Codeine	Morphine	Not suitable
18	G02, G07, G11, G15, G16	Pethidine	Heroin	Not suitable
19	G03, G07, G10, G14, G19	Pethidine	Pethidine	Suitable
20	G08, G11, G15, G18, G19, G20, G22, G23	Codeine	Amphetamines	Not suitable
21	G02, G06, G07, G08, G09, G10, G19, G27	Marijuana	Marijuana	Suitable
22	G05, G11, G12, G14, G19, G20	Hallucinogen	Ecstasy	Not suitable
23	G01, G06, G07, G10, G24	Pethidine	Marijuana	Not suitable
24	G03, G07, G09, G21	Morphine	Morphine	Suitable
25	G05, G12, G17, G22	Ecstasy	Ecstasy	Suitable
26	G07, G12, G15, G26	Heroin	Heroin	Suitable
27	G02, G08, G15, G23	Heroin	Heroin	Suitable
28	G06, G09, G15, G26	Morphine	Morphine	Suitable
29	G07, G11, G18, G23, G27	Codeine	Codeine	Suitable
30	G08, G15, G25, G26	Morphine	Morphine	Suitable

IV. CONCLUSION

The results of this study found that: (a). Machine learning in this study can predict drug users and types of drugs based on the symptoms that drug users complain about (b). This study machine learning acquired knowledge about the symptoms of drug users, types of drugs, and basic knowledge related to the weight of the certainty factor of each type of drug and the symptoms caused so that it can diagnose drug users and the types of drugs used by users. (c). The accuracy of machine learning in this study in predicting the types of drugs used by users and the types of drugs used by users reached 80%. (d). The expert system in this research is website-based so that the expert system from this research can be used by various parties and in different places to identify users and the types of drugs used by users.

The implication of this research result is that the expert system built in this study can be a tool (choice) to replace or complete the testing system for drug users through urine testing in the laboratory.

The drawback of the results of this study is that machine learning expertise in this study is only limited to simple machine learning, as is the case with simple machine learning which was built on previous research by Zhongheng Zhang (2016), which used the KNN method in building learning machines. Furthermore, the machine learning expertise generated from this research is only limited to the expertise possessed in accordance with the knowledge obtained (symptoms, types of drug abuse, rule base, and calculation of certainty factor) under study. Therefore, further research needs to build machine learning that can increase its expertise based on more new data and use another method.

REFERENCES

[1] R. Jiménez, J. Anupol, B. Cajal, and E. Gervilla, "Data mining techniques [illegible]rug use research," *Addict. Behav. Reports*, vol. 8, pp. 128–135, 2018.

[2] [illegible]. Vassoler, E. M. Byrnes, and R. C. Pierce, "The impact of exposure to addictive drugs on future generations: Physiological and behavioral effects," *Neuropharmacology*, vol. 76, no. PART B, pp. 269–275, 2014.

[3] P. K. Shanmugam, "The Influence of Social Factors in Drug Addiction–A Mini Review of Work by Miller & Carroll (2006)," *J. Alcohol. Drug [illegible]pend.*, vol. 05, no. 04, pp. 4–6, 2017.

[4] A. Boys, J. Marsden, and J. Strang, "Understanding reasons for drug use amongst young people: A functional perspective," *Health Educ. Res.*, vol. 16, no. 4, pp. 457–469, 2001.

[5] G. López, L. M. Orchowski, M. K. Reddy, J. Nargiso, and J. E. Johnson, "A review of research-supported group treatments for drug use [illegible]rders," *BMC Public Health*, vol. 16, no. 51, pp. 1–21, 2021.

[6] Z. Justinova, L. V. Panlilio, and S. R. Goldberg, "Drug Addiction," *Natl. Libr. Medicine*, vol. 1, no. 1, pp. 310–335, 2009.

[7] T. Saah, "The evolutionary origins and significance of drug addiction," *Harm Reduct.*, vol. 2, pp. 1–7, 2005.

[8] G. Leshner, M.Stevens, S. Kim, N. Kim, T. L.Wagener, and A. C. Villantie, "Cognitive and affective responses to marijuana prevention and educational messaging," *Drug Alcohol Depend.*, vol. 225, no. August, pp. 1–3, 2021.

[9] M. Zaman *et al.*, "Drug abuse among the students," *Pakistan J. Pharm. Res.*, vol. 1, no. 1, p. 41, 2015.

[10] L. W. Chou, K. M. Chang, and I. Puspitasari, "Drug Abuse Research Trend Investigation with Text Mining," *Comput. Math. Methods Med.*, vol. 2020, pp. 1–8, 2020.

[11] A. Bevan and N. Patel, "An Electronic Prescription Alerting System-Improving the Discharge Medicines Process," *Arch. Dis. Child.*, vol. 101, no. 9, p. e2.55-e2, 2016.

[12] A. Tsyben, N. Gooding, and W. Kelsall, "Assessing the Impact of a Newly Introduced Electronic Prescribing System Across a Paediatric Department – Lessons Learned," *Arch. Dis. Child.*, vol. 101, no. 9, p. 3-e2, 2016.

[13] M. Mahmoudi, S. Pakpour, and G. Perry, "Drug-Abuse Nanotechnology: Opportunities and Challenges," *ACS Chem. Neurosci.*, vol. 9, no. 10, pp. 2288–2298, 2018.

[14] J. Redman, "Recognizing the Warning Signs of Drug Addiction : What You Need to Know," *Mental Health and Counseling Studies*. pp. 1–7, 2021.

[15] N. Jojen, "A Survey Paper on Data Mining Techniques in Drug ustry," *Int. J. Eng. Res. Technol.*, vol. 3, no. 30, pp. 296–299, 2015.

[16] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease gan category," *Front. Chem.*, vol. 6, no. May, pp. 1–11, 2018.

[17] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. nsl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.

[18] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 4, pp. 319–326, 2021.

[19] A. Anggrawan, K. Hidjah, and Q. S. Jihadil, "Kidney failure diagnosis based on case-based reasoning (CBR) method and statistical analysis," in *2016 International Conference on Informatics and Computing, ICIC 16*, 2017, pp. 298–303.

[20] K. Muludi, R. Suharjo, A. Syarif, and F. Ramadhani, "Implementation of forward chaining and certainty factor meth n android-based expert system of tomato diseases identification," *Int. J. Adv. Comput. Sci. pl.*, vol. 9, no. 9, pp. 451–456, 2018.

[21] C. P. C. Munaiseche, D. R. Kaparang, and F. T. D. Rompas, "An Expert System for Diagnosing Eye Diseases using Forward Chaining Method," in *IOP Conference Series: Materials Science and Engineering*, 2018, 306, no. 1, pp. 1–8.

[22] N. Von Greiff and L. Skogens, "Recovery and identity: a five-year follow-up of persons treated in 12-step-related programs," *Drugs Educ. ev. Policy*, pp. 1–10, 2021.

[23] A. Anggrawan, "Percentage of Effect of Blended Learning Madel on Learning Outcome," in *Proceedings of 2019 4th International nference on Informatics and Computing, ICIC 2019*, 2019.

[24] A. Anggrawan, N. Ibrahim, S. Muslim, and C. Satria, "Interaction between Learning Style and Gender in Mixed Learning with 40 % Face-to-face Learning and 60 % Online Learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 5, pp. 407–413, 2019.

Plagiarism check _ camera ready _ Machine Learning for Diagnosing Drug Users and Types of Drugs used

ORIGINALITY REPORT

16%
SIMILARITY INDEX

PRIMARY SOURCES

1	repository.universitasbumigora.ac.id Internet	228 words — 4%
2	www.jatit.org Internet	77 words — 1%
3	Mohammed Jasim A. Alkhafaji, Abbas Fadhil Aljuboori, Abdullahi Abdu Ibrahim. "Clean medical data and predict heart disease", 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2020 Crossref	47 words — 1%
4	journal.uad.ac.id Internet	39 words — 1%
5	docshare.tips Internet	30 words — 1%
6	www.ijrte.org Internet	30 words — 1%
7	Raimondo Maria Pavarin, Dario Consonni. "Early Adolescents and Substance Use", Journal of Addiction, 2013 Crossref	27 words — < 1%

8	Joan Angelina Widians, Novianti Puspitasari, Aidil Febriansyah. "Disease Diagnosis System Using Certainty Factor", 2019 International Conference on Electrical, Electronics and Information Engineering (ICEEIE), 2019 Crossref	26 words — < 1%
9	turcomat.org Internet	26 words — < 1%
10	Arifin Ika Nugroho, Pulung Riyanto, Jori Lahinda, Serli Hatul Hidayat. "Effectiveness of physical condition training on the fitness of drug abuse victims, BRSPP Yogyakarta", Enfermería Clínica, 2020 Crossref	20 words — < 1%
11	www.ijeat.org Internet	20 words — < 1%
12	academic.oup.com Internet	19 words — < 1%
13	www.tandfonline.com Internet	18 words — < 1%
14	Anthony Anggrawan, Nurdin Ibrahim, Suyitno Muslim, Christofer Satria. "Interaction between Learning Style and Gender in Mixed Learning with 40% Face-to-face Learning and 60% Online Learning", International Journal of Advanced Computer Science and Applications, 2019 Crossref	17 words — < 1%
15	docplayer.info Internet	16 words — < 1%
16	Ashok Singh. "Novel nanoparticle-based treatment approaches", Elsevier BV, 2021 Crossref	15 words — < 1%

17	Edgar A. Whitley. "Two approaches to developing expert systems: A consideration of formal and semi-formal domains", AI & Society, 1991 Crossref	15 words — < 1%
18	www.continuingedcourses.net Internet	15 words — < 1%
19	Ania Wajnberg, Katherine Ornstein, Meng Zhang, Kristofer L. Smith, Theresa Soriano. "Symptom Burden in Chronically Ill Homebound Individuals", Journal of the American Geriatrics Society, 2013 Crossref	14 words — < 1%
20	Glenn Leshner, Elise M. Stevens, Amy M. Cohn, Seunghyun Kim, Narae Kim, Theodore L. Wagener, Andrea C. Villanti. "Cognitive and affective responses to marijuana prevention and educational messaging", Drug and Alcohol Dependence, 2021 Crossref	14 words — < 1%
21	W Uriawan, K Manaf, U Syaripudin, M Mujiburrahman. "Expert system for diagnosing narcissistic personality disorders using certainty factor and forward chaining methods", IOP Conference Series: Materials Science and Engineering, 2021 Crossref	14 words — < 1%
22	onlinelibrary.wiley.com Internet	14 words — < 1%
23	Win Min Thit, Sai Wai Yan Myint Thu, Jaranit Kaewkungwal, Ngamphol Soonthornworasiri et al. "User Acceptance of Electronic Medical Record System: Implementation at Marie Stopes International, Myanmar", Healthcare Informatics Research, 2020 Crossref	13 words — < 1%

24	repositorio.unal.edu.co Internet	12 words — < 1%
25	Adekoya Adebayo, Akinwale Adio, Sofoluwe Adetokunbo. "On Algebraic Spectrum of Ontology Evaluation", International Journal of Advanced Computer Science and Applications, 2011 Crossref	11 words — < 1%
26	tropmedhealth.biomedcentral.com Internet	11 words — < 1%
27	Hengki Tamando Sihotang, Fristi Riandari, R.Mahdalena Simanjorang, Agustina Simangunsong, Penda Sudarto Hasugian. "Expert System for Diagnosis Chicken Disease using Bayes Theorem", Journal of Physics: Conference Series, 2019 Crossref	10 words — < 1%
28	Li Ju, Aihua Lyu, Hongxia Hao, Wen Shen, Hua Cui. "Deep Learning-Assisted Three-Dimensional Fluorescence Difference Spectroscopy for Identification and Semiquantification of Illicit Drugs in Biofluids", Analytical Chemistry, 2019 Crossref	9 words — < 1%
29	Zhang, Rui-san, Zhen He, Wei-dong Jin, and Rui Wang. "Effects of the cannabinoid 1 receptor peptide ligands hemopressin, (m)RVD-hemopressin(?) and (m)VD-hemopressin(?) on memory in novel object and object location recognition tasks in normal young and A?1-42-treated mice", Neurobiology of Learning and Memory, 2016. Crossref	9 words — < 1%
30	thesai.org Internet	9 words — < 1%

31	Sutrisno Arianto Pasaribu, Poltak Sihombing, Suherman Suherman. "Expert System for Diagnosing Dental and Mouth Diseases with a Website-Based Certainty Factor (CF) Method", 2020 3rd International Conference on Mechanical, Electronics, Computer, and Industrial Technology (MECnIT), 2020 Crossref	8 words — < 1%
32	unist.dcollection.net Internet	8 words — < 1%
33	www.science.gov Internet	8 words — < 1%
34	Ninive von Greiff, Lisa Skogens. "Recovery and identity: a five-year follow-up of persons treated in 12-step-related programs", Drugs: Education, Prevention and Policy, 2021 Crossref	6 words — < 1%

EXCLUDE QUOTES	OFF	EXCLUDE MATCHES	OFF
EXCLUDE BIBLIOGRAPHY	OFF		

IJACSA Volume 12 Issue 11

VIEW FULL ISSUE

Paper 1: Performance Evaluation of SNMPv1/2c/3 using Different Security Models on Raspberry Pi

Abstract: The Simple Network Management Protocol (SNMP) is one of the dominant protocols for network monitoring and configuration. The first two versions of SNMP (v1 and v2c) use the Community-based Security Model (CSM), where the community is transferred in clear text, resulting in a low level of security. With the release of SNMPv3, the User-based Security Model (USM) and Transport Security Model (TSM) were proposed, with strong authentication and privacy at different levels. The Raspberry Pi family of Single-Board Computers (SBCs) is widely used for many applications. To help their integration into network management systems, it is essential to study the impact of the different versions and security models of SNMP on these SBCs. In this work, we carried out a performance analysis of SNMP agents running in three different Raspberry Pis (Pi Zero W, Pi 3 Model B, and Pi 3 Model B+). Our comparisons are based on the response time, defined as the time required to complete a request/response exchange between a manager and an agent. Since we did not find an adequate tool for our assessments, we developed our own benchmarking tool. We did numerous experiments, varying different parameters such as the type of requests, the number of objects involved per request, the security levels of SNMPv3/USM, the authentication and privacy protocols of SNMPv3/USM, the transport protocols, and the versions and security models of SNMP. Our experiments were executed with Net-SNMP, an open-source and comprehensive distribution of SNMP. Our tests indicate that SNMPv1 and SNMPv2c have similar performance. SNMPv3 has a longer response time, due to the overhead caused by the security services (authentication and privacy). The Pi 3 Model B and Pi 3 Model B+ have comparable performance, and significantly outperform the Pi Zero W.

Author 1: Eric Gamess Author 2: Sergio Hernandez

Keywords: Simple network management protocol; SNMP; performance evaluation; benchmarks; raspberry pi

DOWNLOAD PDF

Paper 2: Digital Economy and its Importance in the Development of Small and Medium Innovative Enterprises

Abstract: A single universally accepted definition, levels and interconnections of digital economy with other economies is not yet developed. Thus, various definitions of the digital economy have been investigated, as well as various approaches to describing the process of transformation of the digital economy for the correct establishment of these relationships. The article observes the relationship between the state of the digital economy, innovative small and medium enterprises, the development of small and medium businesses in general. The stage of transformation of the digital economy of Russia is determined at the second, intermediate stage of development and the main barriers to moving to the third level are pointed out. The dual role of the digital economy in the development of small and medium innovative enterprises is determined based on the selected model of R. Bukht & R. Heeks, the two directions of influence being the SMEs provision with necessary tools and the digital economy becoming the object of innovative development of SMEs. Finally, the assessment of the state of digital economy in Russia is given and the recommendations for its further implementation are given.

Author 1: Tatiana Korsakova Author 2: Lyudmila Dubanevich

Author 3: Oleg Drozdov Author 4: Anna Mikhailova

Author 5: Ekaterina Kamchatova

Keywords: Business; SMEs; entrepreneurship; Russia; digitalization

DOWNLOAD PDF

Paper 3: Human Action Recognition in Video Sequence using Logistic Regression by Features Fusion Approach based on CNN Features

Abstract: Human Action recognition (HAR) gains too much attention due to its wide range of real world applications, such as video surveillance, robotics and computer vision. In video surveillance systems security cameras are placed to monitor activities and motion, generate alerts in undesirable situations. Due to such importance of video surveillance in daily life, HAR becomes the primary and key factor of video surveillance systems. Many researchers worked on human action recognition but HAR still a challenging problem, due to large variation among human to human and human actions in daily life, which make human recognition very challenging and makes surveillance system difficult to outperform. In this article a novel method is proposed by features fusion of pre-trained convolution neural network (CNN) features. Initially pre-trained CNN VGG 19 weights are exploited to extract fully connected 7th layer (FC7) of the selected dataset, subsequently pre-trained fully connected 8th layer features (FC8) extracted by employing pre-trained weights of the same neural network. However the resultant feature fused vector further optimized by employing two

statistical features selection techniques, chi-square test and mutual information to select best features among them to reduced redundancy and increase performance accuracy of human action, a threshold value used for selecting best features. Furthermore the best features are fused, then grid search with 10 fold cross validation is applied for tuning hyper parameter to select best k fold and the resulting best parameter are feed to Logistic regression (LR) classifier for recognition. The proposed technique used You Tube 11 action dataset and achieved 98.49% accuracy. Lastly the proposed method compares with the existing state of the art methods which show dominance performance.

Author 1: Tariq Ahmad Author 2: Jinsong Wu

Author 3: Imran Khan Author 4: Asif Rahim

Author 5: Amjad Khan

Keywords: Human action recognition; logistic regression; deep learning; convolution neural network; features fusion

DOWNLOAD PDF

Paper 4: Application-based Framework for Analysis, Monitoring and Evaluation of National Open Data Portals

Abstract: Open Government Data (OGD) portals are considered of significant national importance towards transparency and accountability improvement. The continuous publication of data in OGD portals introduces the need for high-quality data and the qualitative portal itself. This paper aims to address the data quality issues through a framework composed of several components aimed at measuring and monitoring the OGD portals in an automated way. Through this proposed framework, is intended to monitor and evaluate OGD quality, respectively OGD portals, and to show their progress/regress based on accumulated scores for different periods. The advantage of the proposed framework is the compatibility with any OGD Portal due to its flexibility of integration. The integration interface consists of only a few basic metrics but is necessary that almost the OGD portal possesses and can produce very compressive results. The other advantage is the possibility of extraction of collected data for further analysis and the introduction of artificial intelligence (AI) for prediction purposes to point out how the OGD portals will stand in the next period.

Author 1: Vigan Raca Author 2: Goran Velinov

Author 3: Betim Cico Author 4: Margita Kon-Popovska

Keywords: Open data; government; datasets; evaluation; portals; framework

DOWNLOAD PDF

Paper 5: Deep Learning for Arabic Image Captioning: A Comparative Study of Main Factors and Preprocessing Recommendations

Abstract: Captioning of images has been a major concern for the last decade, with most of the efforts aimed at English captioning. Due to the lack of work done for Arabic, relying on translation as an alternative to creating Arabic

captions will lead to accumulating errors during translation and caption prediction. When working with Arabic datasets, preprocessing is crucial, and handling Arabic morphological features such as Nunation requires additional steps. We tested 32 different variables combinations that affect caption generation, including preprocessing, deep learning techniques (LSTM and GRU), dropout, and features extraction (Inception V3, VGG16). Moreover, our results on the only publicly avail-able Arabic Dataset outperform the best result with BLEU-1=36.5, BLEU-2=21.4, BLEU-3=12 and BLEU4=6.6. As a result of this study, we demonstrated that using Arabic preprocessing and VGG16 image features extraction enhanced Arabic caption quality, but we saw no measurable difference when using Dropout or LSTM instead of GRU.

Author 1: Hani Hejazi Author 2: Khaled Shaalan

Keywords: Deep learning; NLP; Arabic image captioning; Arabic text preprocessing; LSTM; VGG16; INCEPTION V3

DOWNLOAD PDF

Paper 6: Hardware Architecture for Adaptive Dual Threshold Filter and Discrete Wavelet Transform based ECG Signal Denoising

Abstract: The ECG signal, like all signals obtained when instrumenting a data acquisition system, is affected by noises of physiological and technical sources such as Electromyogram (EMG) and power line interferences, which can deteriorate its morphology. To overcome this issue, it’s subjected to apply a preprocessing step to remove these noises. Filtring techniques are complex computations becoming more common in medical applications, which must be completed in real-time. As a result, these applications are geared at integrating high-performance embedded architectures. This paper presents an FPGA (Field Programmable Gate Array) embedded architecture designed for an ECG denoising hybrid technique based on the Discrete Wavelet transform (DWT) and the Adaptive Dual Threshold Filter (ADTF), dedicated to handle with noises affecting ECG signals. The architecture was designed following a hardware-software codesign using a high-level description language and synthetized to be implemented on different FPGAs due to the structural description flexibility. The global architecture was divided into a set of functional blocks to allow parallel processing of ECG data. The simulation results confirm the high performance of the system in noise reduction without affecting the morphology of the signal. The process takes 0.3 ms with an acquisition frequency of 360 Hz. The whole architecture requires a small area in different FPGAs in terms of resources utilization. It uses less than 1% of the total registers for all FPGA devices which represents a total of 292 registers for Cyclone III LS, Cyclone IV GX, Cyclone IV E, and Arria II GX; and a total of 329 registers for Cyclone V. The logic elements occupancy varies between 3% using Cyclone V and 60% using Cyclone IV GX freeing up space for other parallel processing tasks.

Author 1: Safa MEJHOUDI

Author 2: Rachid LATIF Wissam JENKAL

Author 3: Amine Saddik

Author 4: Abdelhafid EL OUARDI SATIE

Keywords: ECG signal; DWT; ADTF; hybrid technique; hardware-software codesign; FPGA

DOWNLOAD PDF

Paper 7: Machine Learning based Forecasting Systems for Worldwide International Tourists Arrival

Abstract: The international tourist movement has overgrown in recent decades, and travelers are considered a significant source of income to the tourism economy. When tourists visit a place, they spend considerable money on their enjoyment, travel, and hotel accommodations. In this research, tourist data from 2010 to 2020 have been extracted and extended with depth analysis of different dimensions to identify valuable features. This research attempts to use machine learning regression techniques such as Support Vector Regression (SVR) and Random Forest Regression (RFR) to forecast and predict worldwide international tourist arrivals and achieved forecasting accuracy using SVR is 99.4% and using RFR is 84.7%. The study also analyzed the forecasting deadlock condition after covid-19 in the sudden drop of international visitors due to lockdown enforcement by all countries.

Author 1: Ram Krishn Mishra
Author 2: Siddhaling Urolagin
Author 3: J. Angel Arul Jothi
Author 4: Nishad Nawaz
Author 5: Haywantee Ramkissoon

Keywords: Tourists; forecasting; machine learning; Covid-19

DOWNLOAD PDF

Paper 8: Analyzing the Sentiments of Jordanian Students Towards Online Education in the Higher Education Institutions

Abstract: Sentiment analysis and opinion polling are two areas that have grown significantly over the past decade. Opinion research and sentiments analysis in the online education environment can truly reflect the learning state of students and educators and experts in the field; providing the theoretical basis needed to further review educational procedure and conduct. This study aims to shed light on identifying and visualizing students' objective feelings based on an exploration of the subject matter and materials of learning and gathering sentiments from university Facebook groups at various levels and layers in detail. The proposed method is a qualitative descriptive research method that includes data pre-processing, subject discovery, sentiment analysis, and visualization. In relative terms, 39.7% of text messages were positive and 52.3% of text messages were negative and understanding the narrative of these feelings and their impact on the online learning environment.

Author 1: Bayan Alfayoumi
Author 2: Mohammad Alshraideh
Author 3: Saleh Al-Sharaeh
Author 4: Martin Leiner
Author 5: Iyad Muhsen AlDajani

Keywords: Online education; students; sentiment analysis; online education; online environment; online social media

DOWNLOAD PDF

Paper 9: Comparative Study of Flooding Area Detection with SAR Images based on Thresholding and Difference Images Acquired Before and After the Flooding

Abstract: Comparative study of flooding area detection with Synthetic Aperture Radar (SAR) images based on thresholding and difference images acquired before and after the flooding is conducted. Method for flooding, landslide and sediment disaster area detections with SAR is proposed. The following two different methods for flooding detection are common. It is not so easy to determine a threshold for the thresholding method while subtraction method between before and after images of a disaster occurrence has the disadvantage that false disaster areas are detected due to a variation of ground cover targets. Therefore, a comparative study between both methods is required. Its application is demonstrated for the disaster which is occurred in Saga Prefecture, Japan due to a long term of heavy rain during from the begging of August to the middle of August in 2021. Through experiments with Sentinel-1 SAR imagery data, it is found that the proposed method works well for the detection of the disaster.

Author 1: Kohei Arai

Keywords: Flooding; landslide; sediment disaster; heavy rain; image quality; Synthetic Aperture Radar; SAR; sentinel-1 SAR; thresholding; difference images between before and after disaster

DOWNLOAD PDF

Paper 10: Developing of Middleware and Cross Platform Chat Application

Abstract: The rapid development of technology has resulted in many new innovations on social media platforms. Now-a-days, there are many chat applications available, namely Whatsapp, Telegram, LINE, Viber, and many others. This in turn forces users to juggle between many chat applications as different applications can’t communicate with each other. This research aims to develop a chat application which serves as a middleware to make communication between developed chat application and two conventional chat applications possible (Telegram and LINE). Several tests are done to ensure that the message exchange process (in text, picture, video, and file type) works well between the developed chat application as well as Telegram or LINE.

Author 1: Danny Sebastian Author 2: Restyandito

Author 3: Kristian Adi Nugraha

Keywords: Telegram API; line API; chat application; flutter; middleware

DOWNLOAD PDF

Paper 11: A Survey on Deep Learning Face Age Estimation Model: Method and Ethnicity

Abstract: Face age estimation is a type of study in computer vision and pattern recognition. Designing an age estimation or classification model requires data as training samples for the machine to learn. Deep learning method has improved estimation accuracy and the number of deep learning age estimation models developed. Furthermore, numerous datasets availability is making the method an increasingly attractive approach. However, face age databases mostly have limited ethnic subjects, only one or two ethnicities and may result in ethnic bias during age estimation, thus impeding progress in understanding face age estimation. This paper reviewed available face age databases, deep learning age estimation models, and discussed issues related to ethnicity when estimating age. The review revealed changes in deep learning architectural designs from 2015 to 2020, frequently used face databases, and the number of different ethnicities considered. Although model performance has improved, the widespread use of specific few multi-races databases, such as the MORPH and FG-NET databases, suggests that most age estimation studies are biased against non-Caucasians/non-white subjects. Two primary reasons for face age research's failure to further discover and understand ethnic traits effects on a person's facial aging process: lack of multi-race databases and ethnic traits exclusion. Additionally, this study presented a framework for accounting ethnic in face age estimation research and several suggestions on collecting and expanding multi-race databases. The given framework and suggestions are also applicable for other secondary factors (e.g. gender) that affect face age progression and may help further improve future face age estimation research.

Author 1: Hadi A. Dahlan

Keywords: Deep learning; face age estimation; face database; ethnicity bias

DOWNLOAD PDF

Paper 12: Comparative Analysis of Supervised Machine Learning Techniques for Sales Forecasting

Abstract: This study talks about how data mining can be used for sales forecasting in retail sales and demand prediction. Prediction of sales is a crucial task which determines the success of any organization in the long run. There are various techniques available for predicting the sales of a supermarket such as Time Series Algorithm, Regression Techniques, Association rule etc. In this paper, a comparative analysis of some of the Supervised Machine Learning Techniques have been done such as Multiple Linear Regression Algorithm, Random Forest Regression Algorithm, K-NN Algorithm, Support Vector Machine (SVM) Algorithm and Extra Tree Regression to build a prediction model and precisely estimate possible sales of 45 retail outlets of Walmart store which are at different geographical locations. Walmart is one of the foremost stores across the world and thus authors would like to predict the sales accurately. Certain events and holidays affect the sales periodically, which sometimes can also be on a daily basis. The forecast of probable sales is based on a combination of features such as previous sales data, promotional events, holiday week, temperature, fuel price, CPI i.e., Consumer Price Index and Unemployment rate in the state. The data is collected from 45 outlets of Walmart and the prediction about the sales of Walmart was done using various Supervised Machine Learning Techniques. The contribution of this paper is to help the business owners decide which approach to follow while trying to predict the sales of their Supermarket taken into account different scenarios including temperature, holidays, fuel price, etc. This will help them in deciding the promotional and marketing strategy for their products.

Author 1: Stuti Raizada Author 2: Jatinderkumar R. Saini

Keywords: Sales forecasting; linear regression; random forest regression; KNN regression algorithm; SVM algorithm; supervised machine learning techniques

DOWNLOAD PDF

Paper 13: Machine Learning for Diagnosing Drug Users and Types of Drugs Used

Abstract: Drug use is very detrimental to the physical and psychological health of users. Drug abuse also causes addiction and is a global epidemic. Therefore it is not surprising that scientific research related to drugs has attracted attention for research. However, many factors become obstacles in the medical services of the drug user, including cost, flexibility, and a slow process. Meanwhile, electronic systems can speed up handling time, improve work efficiency, save costs and reduce inspection errors. It means that a breakthrough is needed in developing a platform that can identify drug users. Therefore, this research aims to build machine learning with expertise like an expert who can diagnose drug users and distinguish the types of drugs used by drug users. The expert system on machine learning was developed using the Forward Chaining and Certainty Factor methods. This study concludes that the expert system on machine learning developed can be used to diagnose drug users and distinguish the types of drugs used with an accuracy of up to 80%. The implications of the expert system on machine learning are an alternative method for narcotics officers and medical doctors in diagnosing drug users and the types of drugs used.

Author 1: Anthony Anggrawan Author 2: Christofer Satria

Author 3: Che Ku Nuraini Author 4: Lusiana

Author 5: Ni Gusti Ayu Dasriani Author 6: Mayadi

Keywords: Machine learning; drug; expert system; forward chaining; certainty factor

DOWNLOAD PDF

Paper 14: Insights on Deep Learning based Segmentation Schemes Towards Analyzing Satellite Imageries

Abstract: Satellite imageries are essentially a complex form of an image when subjected to critical analytical operation. The analytical process applied on remotely sensed satellite imageries are utilized for generating the land cover map. With an abundance of traditional techniques evolved to date, deep learning-based schemes are progressively gaining pace for identifying and classifying a terrestrial object in satellite images. However, different variants of deep learning approaches have different operations, and so are the consequences. At the same time, there is no reported literature to highlight the issues, trends, and effectiveness much on a generalized scale concerning segmentation. Therefore, this paper reviews some of the recent segmentation approaches using deep learning to contribute towards review findings in the form of research trends, research gaps, and essential learning outcomes. The

paper offers a compact and distinct picture of deep learning approaches used to boost segmentation for satellite images.

Author 1: Natya S Author 2: Ramya K

Author 3: Seema Singh

Keywords: Deep learning; landcover; map generation; remotely sense image; satellite image; segmentation

DOWNLOAD PDF

Paper 15: DNA Profiling: An Investigation of Six Machine Learning Algorithms for Estimating the Number of Contributors in DNA Mixtures

Abstract: DNA (Deoxyribonucleic acid) profiling involves analysis of sequences of individual or mixed DNA profiles to identify persons these profiles belong to. DNA profiling is used in important applications such as for paternity tests, in forensic science for person identification on a crime scheme, etc. Finding the number of contributors in a DNA mixture is a major task in DNA profiling with challenges caused due to allele dropout, stutter, blobs, and noise. The existing methods for finding the number of unknowns in a DNA mixture suffer from issues including computational complexity and accuracy of estimating the number of unknowns. Machine learning has received attention recently in this area but with limited success. Many more efforts are needed for improving the robustness and accuracy of these methods. Our research aims to advance the state-of-the-art in this area. Specifically, in this paper, we investigate the performance of six machine learning algorithms -- Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR), Stochastic Gradient Descent (SGD), and Gaussian Naïve-Bayes (GNB) -- applied to a publicly available dataset called PROVEDIt, containing mixtures with up to five contributors. We evaluate the algorithmic performance using confusion matrices and four performance metrics namely accuracy, F1-Score, Recall, and Precision. The results show that LR provides the highest Accuracy of 95% for mixtures with five contributors.

Author 1: Hamdah Alotaibi Author 2: Rashid Mehmood

Author 3: Fawaz Alsolami

Keywords: Machine learning; DNA profiling; DNA mixtures; forensic science

DOWNLOAD PDF

Paper 16: Visualization of the Temporal Topic Model on Higher Education Preferences with Higher Education Ranking Indicators

Abstract: Private universities have devised a strategy to counteract the ongoing competition. Private universities can use the appropriate data analysis method to make higher education management decisions. The goal of this research is to find a new approach to data analysis methods in the form of visualization using the TTM (Temporal Topic Model) method to assist private university management. These findings are the two formulas used to generate time-based visualizations and the Temporal Topic Model per month to visually

change news topics related to rankings so that management can decide on marketing strategies and policies that are in relation to public opinion.

Author 1: Winda Widya Ariestya

Author 2: Achmad Benny Mutiara Author 3: I Made Wiryana

Author 4: Setia Wirawan

Keywords: Management decisions; temporal topic model; university; visualization

DOWNLOAD PDF

Paper 17: Sparse Distributed Memory Approach for Reinforcement Learning Driven Efficient Routing in Mobile Wireless Network System

Abstract: In recent years, researchers have explored the applicability of Q-learning, a model-free reinforcement learning technology towards designing QoS-aware, resource-efficiency, and reliable routing techniquesin a dynamically changing network environment. However, Q-learning is based on tabular representation to characterize learned policies that frequently encounter a dimension disaster problem when introduced to the uncertain and dynamically changing network environment. In addition, the time required for agent learning in the training phase is too long, which makes it difficult for the agent to generalize the observation state efficiently. To this end, this paper attempts to overcome the overhead memory problems encountered in Q-learning-based routing techniques. In this paper, the study presents a novel memory-efficient intelligent routing mechanism based on adaptive Kanerva coding, which minimizes the storage cost required for storing large action and a state value. Unlike existing schemes, the proposed method optimizes memory requirements. Also, it enables better generalization by storing the learnable parameters of the function approximator present in the agent in a Kanerva-coding data structure. The Kanerva-coding is a sparse memory with distributed reading and writing mechanism which enables optimal compression and state abstractions for learning with fewer parameterized components making it highly memory efficient. The design and implementation of the proposed technique are done on the Anaconda tool. Simulation results demonstrate that the proposed technique can adaptively adjust the routing policy according to the varying network environment to meet the transmission requirements of different services with low memory requirements.

Author 1: Varshini Vidyadhar Author 2: Nagaraj R

Author 3: G Sudha

Keywords: Mobile wireless network; reinforcement learning; Q-learning; Kanerva coding; routing; memory optimization

DOWNLOAD PDF

Paper 18: A Systematic Review of Published Articles, Phases and Activities in an Online Social Networks Forensic Investigation Domain

Abstract: The purpose of this paper is to retrieve, evaluate and analyse the available published articles in five (5) relevant online databases from 2011 to 2021 and also critically identify the phases and activities involved in an Online Social Networks Forensic Investigation based on bibliometric analysis and Degree of confidence respectively in order to know the evolution in the research domain. A systematic literature review (SLR) technique was adopted by the author to search using pre-defined keywords. Only scholarly articles published between 2011 and 2021 written in English were included in the search. The total of 316 subscribed documents were collected from the five (5) online databases based on the search criteria although twenty-nine (29) are duplicates. ScienceDirect has the highest number with 189 documents and the year 2020 with the highest published articles. Six (6) phases and forty-three (43) activities were identified. According to a review of the recovered publications, no previous research has been done to statistically retrieve, evaluate and analyse the level of work that has been done in the domain of OSNFI, as well as the phases and activities involved in the forensic investigation of an online social networks crime.

Author 1: Aliyu Musa Bade Author 2: Siti Hajar Othman

Keywords: Forensic; investigation; model; online social networks; bibliometric analysis; degree of confidence

DOWNLOAD PDF

Paper 19: Deep Learning based Neck Models for Object Detection: A Review and a Benchmarking Study

Abstract: Artificial intelligence is the science of enabling computers to act without being further programmed. Particularly, computer vision is one of its innovative fields that manages how computers acquire comprehension from videos and images. In the previous decades, computer vision has been involved in many fields such as self-driving cars, efficient information retrieval, effective surveillance, and a better understanding of human behaviour. Based on deep neural networks, object detection is actively growing for pushing the limits of detection accuracy and speed. Object Detection aims to locate each object instance and assign a class to it in an image or a video sequence. Object detectors are usually provided with a backbone network designed for feature extractors, a neck model for feature aggregation, and finally a head for prediction. Neck models, which are the purpose of study in this paper, are neural networks used to make a fusion between high-level features and low-level features and are known by their efficiency in object detection. The aim of this study to present a review of neck models together before making a benchmarking that would help researchers and scientists use it as a guideline for their works.

Author 1: Sara Bouraya Author 2: Abdessamad Belangour

Keywords: Object detection; deep learning; computer vision; neck models; feature aggregation; feature fusion

DOWNLOAD PDF

Paper 20: Naïve Bayes Classification of High-Resolution Aerial Imagery

Abstract: In this study, the performance of Naïve Bayes classification on a high-resolution aerial image captured from a UAV-based remote sensing platform is investigated. K-means clustering of the study area is initially performed to assist in selecting the training pixels for the Naïve Bayes classification. The Naïve Bayes classification is performed using linear and quadratic discriminant analyses and by making use of training set sizes that are varied from 10 through 100 pixels. The results show that the 20 training set size gives the highest overall classification accuracy and Kappa coefficient for both discriminant analysis types. The linear discriminant analysis with 94.44% overall classification accuracy and 0.9395 Kappa coefficient is found higher than the quadratic discriminant analysis with 88.89% overall classification accuracy and 0.875 Kappa coefficient. Further investigations carried out on the producer accuracy and area size of individual classes show that the linear discriminant analysis produces a more realistic classification compared to the quadratic discriminant analysis particularly due to limited homogenous training pixels of certain objects.

Author 1: Asmala Ahmad Author 2: Hamzah Sakidin

Author 3: Mohd Yazid Abu Sari

Author 4: Abd Rahman Mat Amin

Author 5: Suliadi Firdaus Sufahani Author 6: Abd Wahid Rasib

Keywords: Naïve Bayes; k-means; classification accuracy; training set size; discriminant analysis

DOWNLOAD PDF

Paper 21: Secured and Provisioned Access Authentication using Subscribed user Identity in Federated Clouds

Abstract: Cloud computing has become an essential source for modern trade or market environments by abled frameworks. The exponential growth of cloud computing services in the last few years has resulted in extensive use, especially in storing and sharing the data on various cloud servers. The current trend in the cloud shows that the cloud owners use relative functions and target areas in such a way that cloud customers access or store their data either in the same servers or related servers. Simultaneously, from the security point of view, the lack of confidence about the customer's data on the cloud server is still questionable. The hour's need is to provide the cloud service in a 'single port way' by forming the joint management policy to increase customer satisfaction and profitability. In addition to this, the authentication steps also need to be improvised. This paper discusses issues on the security authentication and access provisioning of the cloud service consumers in federated clouds using subscribed user identity. This work proposes the user identity verification module (UidVM) in the cloud service consumer's authentication process to serve as a cloud broker to minimize the work overloads on the central cloud federation management system, thus enhancing the cloud security.

Author 1: Sudan Jha Author 2: Sultan Ahmad

Author 3: Meshal Alharbi Author 4: Bader Alouffi

Author 5: Shoney Sebastian

Keywords: Security authentication (SA); cloud federation (CF); cloud service provider (CSP); key distribution center (KDC); user identity verification module (UIdVM)

DOWNLOAD PDF

Paper 22: Local Frequency Descriptor and Hybrid Features for Classification of Brain Magnetic Resonance Images using Ensemble Classifier

Abstract: A brain tumor is an irregular development of cells in the human brain that causes problems with the brain's normal functionalities. Early detection of brain tumor is an essential process to help the patient to live longer than treatment. Hence in this paper, a hybrid ensemble model has been proposed to classify the input brain MRI images into two classes: brain MRI images having tumor and brain MRI images with no tumor. The hybrid features are extracted by analyzing the texture and statistical properties of brain MRI images. Further, the Local Frequency Descriptor (LFD) technique is employed to extract the prominent features from the brain tumor region. Finally, an ensemble classifier has been developed with the combination of Support Vector Machine (SVM), Decision Tree (DT) and K-Nearest Neighbour (KNN) technique to successfully classify the brain MRI images into brain tumor MRI images and non-tumor brain MRI images. The proposed model is tested on the Kaggle brain tumor dataset and the performance of the method is evaluated in terms of accuracy, sensitivity, specificity, precision, recall and f-measure (f1 score-harmonic mean of precision and recall). The results show that the proposed model is promising and encouraging.

Author 1: Shruthi G Author 2: Krishna Raj P M

Keywords: Brain tumor; hybrid features; local frequency descriptor (LFD); ensemble classifier

DOWNLOAD PDF

Paper 23: Mutual Informative Brown Clustering based Multiattribute Cockroach Swarm Optimization for Reliable Data Dissemination in VANET

Abstract: A vehicular ad hoc network (VANETs) intends to obtain communication for vehicular networks and enhances road safety and effectiveness with help of wireless technology. Data dissemination is an important process in communication. In VANETs system, Data dissemination plays a significant role. A novel Mutual informative brown clustering-based multi attribute cockroach swarm optimization (MIBCMCSO) technique is introduced for improving data dissemination. In reliable data dissemination, clustering and optimization are the two major process of proposed MIBCMCSO technique. Initially, clustering procedure is performed for separating entire network towards different groups of vehicle nodes namely distance, direction, density and velocity of node. For each group, cluster head was chosen among the members to efficient data with minimum delay. Secondly, multi attribute cockroach swarm optimization technique is applied for finding optimal cluster head through multi attribute functions such as residual energy, bandwidth availability, and distance. Then source node performs data dissemination destination via optimal cluster head. Simulation of MIBCMCSO as well as existing technique is performed by various performance parameters like packet

delivery ratio, end to end delay and throughput. MIBCMCSO achieves higher consistency of data dissemination as well as lesser delay than conventional methods.

Author 1: D. Radhika Author 2: A. Bhuvaneswari

Keywords: VANET; data dissemination; mutual informated brown agglomerative clustering; multi attribute cockroach swarm optimization

DOWNLOAD PDF

Paper 24: Face Age Estimation and the Other-race Effect

Abstract: Age estimation is an automated method of predicting human age from 2-D facial feature representations. The majority of studies carried out in this research area use the FG-NET and MORPH 2 databases to train and test developed systems, which are lacking in black-face content. Most age frauds are perpetuated in the sub-Saharan African region due to the unavailability of an official database and unregistered births in the rural areas. The issues of unverified age in the region made it possible for under-age voters, under-age drivers, and the engagement of over-aged sportsmen. The other-race effect could reduce the performance of face recognition techniques, which could make techniques that work for white faces underperform when deployed for use in the predominately black face region. This study examines the other-race effect on face-based age estimation by analyzing the accuracy of an age estimation system trained with predominantly black faces against the same age estimation system trained with predominately white faces. The developed age estimation system uses a genetic algorithm-artificial neural network classifier and local binary pattern for texture and shape feature extraction. A total of 170 black faces were used for system testing. The result showed that the age estimation system trained with the predominantly black face database (GA-ANN-AES-855) outperformed the system trained with predominantly white faces (GA-ANN-AES-255) on testing with the aforementioned black face samples. The results obtained from the simulation were further subjected to inferential statistics, which established that the improvement in the correct classification rate was statistically significant. Hence, the other-race effect affects face-based age estimation systems.

Author 1: Oluwasegun Oladipo

Author 2: Elijah Olusayo Omidiora

Author 3: Victor Chukwudi Osamor

Keywords: Component; face recognition; age estimation; other-race effect

DOWNLOAD PDF

Paper 25: Polarity Detection of Dialectal Arabic using Deep Learning Models

Abstract: With the evolution of a new era of technology and social media networks, as well as an increase in Arabs sharing their point of view, it became necessary that this research be conducted. Sentiment analysis is concerned with identifying and extracting opinionated phrases from reviews or tweets. Specifically, to determine whether a given tweet is positive, negative, or

neutral. Dialectical Arabic poses difficulties for sentiment analysis. In this paper, four deep learning models are presented, to be specific convolution neural networks (CNN), long short-term memory (LSTM), a hybrid of (CNN-LSTM), and Bidirectional LSTMs (BiLSTM), to determine the tweets polarities written in dialectal Arabic. The performance of the four models is validated on the used corpus with the use of word embedding and applying the (k-Fold Cross-Validation) method. The results show that CNN outperforms the others achieving an accuracy of 99.65%.

Author 1: Saleh M. Mohamed

Author 2: Ensaf Hussein Mohamed

Author 3: Mohamed A. Belal

Keywords: Sentiment analysis; word embedding; sentiment classification; dialectical arabic; deep learning

DOWNLOAD PDF

Paper 26: Challenges in Developing Virtual Reality, Augmented Reality and Mixed-Reality Applications: Case Studies on A 3D-Based Tangible Cultural Heritage Conservation

Abstract: A model that contributes in a simple, practical and effective way to develop 3D-based CH conservation applications involving the use of VR, AR and MR technologies was proposed based on the identification of challenges in developing applications. Identification was carried out by analyzing related and relevant articles selected randomly using Google and Google Scholar search engines. The model can prevent researchers from lack of planning in carrying out research in this field, and it is suitable for those just starting out with this type of research. In addition, this model can support researchers to more easily, practically and effectively implement 3D-based cultural heritage conservation by using virtual reality or augmented reality or mixed reality technology.

Author 1: Ahmad Zainul Fanani Author 2: Khafiizh Hastuti

Author 3: Arry Maulana Syarif

Author 4: Prayanto Widyo Harsanto

Keywords: Virtual reality; augmented reality; mixed reality; tangible cultural heritage; 3D-based cultural heritage conservation

DOWNLOAD PDF

Paper 27: Trust-based Key Management Conglomerate ElGamal Encryption for Data Aggregation Framework in WSN using Blockchain Technology

Abstract: In wireless sensor networks (WSN), data aggregation is a widely used method. Security issues like data integrity and data confidentiality became a significant concern in data aggregation when the sensor network is deployed in a hostile environment. Many researches may carry out several works to tolerate these security issues. However, there were some limitations like delay, the arrival rate of packets, and so on. Hence, to overcome the

existing problems, this approach offers a blockchain-dependent data aggregation scheme in WSN. The main intention of the proposed work is to generate a certificateless key generation so that the proposed system's secrecy rate is improved. The use of blockchain is employed for security purposes, and it enables the user to acquire the information stored internally in an effortless manner. Initially, deployment of sensor and base station (BS) is carried out, followed by node registration at which the public/private keys are generated. The computation of private hash values is carried by performing certificateless key generation. After that, the formation of blockchain is made using the PoW (Proof of Work) detection algorithm followed by the aggregation of data. In the data aggregation process, Elgammal based cryptographic approach is introduced to acquire member data, perform aggregation logic, and transfer the aggregated data. Finally, cluster-based routing is established with the use of Knapsack based cluster routing strategy. The performance investigation of the proposed system is estimated and the outcomes attained are compared with the existing techniques in terms of arrival rate, average delay, and the delay ratio of the packets. The investigation illustrates that the suggested approach is better than the traditional techniques.

Author 1: T. G. Babu Author 2: V. Jayalakshmi

Keywords: Wireless sensor networks; data aggregation; PoW detection scheme; blockchain technology; cluster formation; key generation; security; delay ratio

DOWNLOAD PDF

Paper 28: Statistical Analysis for Revealing Defects in Software Projects: Systematic Literature Review

Abstract: Defect detection in software is the procedure to identify parts of software that may comprise defects. Software companies always seek to improve the performance of software projects in terms of quality and efficiency. They also seek to deliver the soft-ware projects without any defects to the communities and just in time. The early revelation of defects in software projects is also tried to avoid failure of those projects, save costs, team effort, and time. Therefore, these companies need to build an intelligent model capable of detecting software defects accurately and efficiently. The paper is organized as follows. Section 2 presents the materials and methods, PRISMA, search questions, and search strategy. Section 3 presents the results with an analysis, and discussion, visualizing analysis and analysis per topic. Section 4 presents the methodology. Finally, in Section 5, the conclusion is discussed. The search string was applied to all electronic repositories looking for papers published between 2015 and 2021, which resulted in 627 publications. The results focused on finding three important points by linking the results of manuscript analysis and linking them to the results of the bibliometric analysis. First, the results showed that the number of defects and the number of lines of code are among the most important factors used in revealing software defects. Second, neural networks and regression analysis are among the most important smart and statistical methods used for this purpose. Finally, the accuracy metric and the error rate are among the most important metrics used in comparisons between the efficiency of statistical and intelligent models.

Author 1: Alia Nabil Mahmoud Author 2: Vítor Santos

Keywords: Defects; software projects; statistical model; linear regression; logistic regression

DOWNLOAD PDF

Paper 29: A Hybrid Deep Neural Network for Human Activity Recognition based on IoT Sensors

Abstract: Internet of things (IOT) sensors, has received a lot of interest in recent years due to the rise of application demands in domains like ubiquitous and context-aware computing, activity surveillance, ambient assistive living and more specifically in Human activity recognition. The recent development in deep learning allows to extract high-level features automatically, and eliminates the reliance on traditional machine learning techniques, which depended heavily on hand crafted features. In this paper, we introduce a network that can identify a variety of everyday human actions that can be carried out in a smart home environment, by using raw signals generated from Internet of Thing's motion sensors. We design our architecture basing on a combination of convolutional neural network (CNN) and Gated recurrent unit (GRU) layers. The CNN is first deployed to extract local and scale-invariance features, then the GRU layers are used to extract sequential temporal dependencies. We tested our model called (CNGRU) on three public datasets. It achieves an accuracy better or comparable to existing state of the art models.

Author 1: Zakaria BENHAILI Author 2: Youssef BALOUKI

Author 3: Lahcen MOUMOUN

Keywords: IoT; deep learning; CNN; GRU; CNGRU; human activity recognition

DOWNLOAD PDF

Paper 30: Bioinformatics Research Through Image Processing of Histopathological Response to Stonefish Venom

Abstract: The present study utilizes coastal and environmental engineering to investigate the histopathological effects of Synanceia verrucosa venom on Albino BALB/c mice. S. verrucosa is the most hazardous venomous marine fish that belong to the family Synanceiidae, generally known as the "Reef Stonefish". Crude venom was collected from venom glands of the dorsal spines of stonefish samples taken from the Jordanian coastline of the Gulf of Aqaba, Red Sea. The mice were given intramuscular injections of the venom. Consequently, the research evaluated the acute toxicity and influence on selected serum biomarker enzymes, as well as possible histological alterations of the soleus skeletal muscles. The mice 24 h LD50 was 0.107 µg toxin/kg mouse body weight. After the treatment using venom sublethal dose, the serum biomarkers, including Lactate dehydrogenase (LDH) and Alanine aminotransferase (ALT), were significantly improved ($P \leq 0.05$). In addition, Lipid Peroxidation (LPO) contents were significantly increased ($P \leq 0.05$) after venom treatment. Moreover, we combined routine medical procedures and artificial intelligence-assisted image analysis for a rapid qualitative and quantitative diagnosis of stonefish injury, based on the histophotography of mice tissue samples during the observation period (1, 2, and 3 hours respectively). The novelty of our method is that we could detect severe and mild damage with an accuracy of 93% and 91%, respectively. The most histological abnormalities in muscles were the great variety in diameters, content, and widespread among

randomly distributed muscle fibres. In addition, loss of the tissue's striated appearance was noticed in toxin-treated groups compared with the control group. Consequently, our findings indicate the Stonefish's harmful influences that may endanger human life and highlight the need for appropriate measures to be considered. This, in turn, can ensure beach safety in the Gulf of Aqaba.

Author 1: Mohammad Wahsha Author 2: Heider A. M. Wahsheh

Author 3: Wissam Hayek Author 4: Haya Al-Tarawneh

Author 5: Maroof Khalaf Author 6: Tariq Al-Najjar

Keywords: Synanceia verrucosa; Gulf of Aqaba; artificial intelligence; marine biotoxins

DOWNLOAD PDF

Paper 31: Real Time Distributed and Decentralized Peer-to-Peer Protocol for Swarm Robots

Abstract: This contribution proposes an approach to enhance the capability of robotic agents to join the Internet of Things (IoT) and act autonomously in extreme and hostile environment. This capability will help in the development in environments where the connectivity, availability, and responsivity of the devices are subject to variations and noises. A real time distributed and decentralized Peer-to-Peer protocol was designed to allow Autonomous Unmanned Surface Vessels (AUSV) extend their context awareness. The developed Middleware allows a real time communication and is designed to run on top of a Real Time Operating System (RTOS). Furthermore, the proposed Middleware will give researchers access to a large amount of data collected by sensors, and thus solve one of the major problems encountered while training artificial intelligence models which is the lack of sufficient data.

Author 1: Mahmoud Almostafa RABBAH

Author 2: Nabila RABBAH Author 3: Hicham BELHADAOUI

Author 4: Mounir RIFI

Keywords: Autonomous robots; smart objects; peer-to-peer; real time communication; ROS2; ZeroMQ; middleware

DOWNLOAD PDF

Paper 32: Improving Customer Churn Classification with Ensemble Stacking Method

Abstract: Due to the high cost of acquiring new customers, accurate customer churn classification is critical in any company. The telecommunications industry has employed single classifiers to classify customer churn; however, the classification accuracy remains low. Nevertheless, combining several classifiers' decisions improves classification accuracy. This article attempts to enhance ensemble integration via stack generalisation. This paper proposed a stacking ensemble based on six different learning algorithms as the base-classifiers and tested on five different meta-model classifiers. We compared the performance of the proposed stacking ensemble model with single classifiers, bagging and boosting ensemble. The performances of the models

were evaluated with accuracy, precision, recall and ROC criteria. The findings of the experiments demonstrated that the proposed stacking ensemble model resulted in the improvement of the customer churn classification. Based on the results of the experiments, it indicates that the prediction accuracy, precision, recall and ROC of the proposed stacking ensemble with MLP meta-model outperformed other single classifiers and ensemble methods for the customer churn dataset.

Author 1: Mohd Khalid Awang Author 2: Mokhairi Makhtar

Author 3: Norlina Udin Author 4: Nur Farraliza Mansor

Keywords: Stacking ensemble; customer churn prediction; bagging; boosting

DOWNLOAD PDF

Paper 33: Enhancing the Takhrij Al-Hadith based on Contextual Similarity using BERT Embeddings

Abstract: Muslims are required to conduct Takhrij to validate the truth of Hadith text, especially when it is obtained from online media. Typically, the traditional Takhrij processes are conducted by experts and apply to Arabic Hadith text. This study introduces a contextual similarity model based on BERT Embedding to handle Takhrij on Indonesian Hadith Text. This study examines the effectiveness of BERT Fine-Tuning on the six pre-trained models to produce embedding models. The result shows that BERT Fine-Tuning improves the embedding model average accuracy by 47.67%, with a mean of 0.956845. The most high-grade accuracy was the BERT embedding built based on the indobenchmark/indobert-large-p2 pre-trained model on 1.00. In addition, the manual evaluation achieved 91.67% accuracy.

Author 1: Emha Taufiq Luthfi

Author 2: Zeratul Izzah Mohd Yusoh

Author 3: Burhanuddin Mohd Aboobaider

Keywords: Hadith text; Takhrij; natural language processing; text-similarity; word embedding; BERT fine-tuning

DOWNLOAD PDF

Paper 34: UX Testing for Mobile Learning Applications of Deaf Children

Abstract: Many studies are focusing on deaf children mobile learning. However, they are not concentrating on user experience (UX) testing. Current UX testing is based on existing UX evaluation models that are hard to apply due to the comprehensive measurements and lack of description on how to conduct evaluation for a more specific mobile learning process. Moreover, the existing UX evaluation models are not highlighted to be applied in testing UX for deaf children's mobile learning. Hence, this paper proposed questions for UX testing for deaf children's mobile learning to explore UX issues in offering an enjoyable learning application. Smileyometer is used to capture the data from deaf children after using a selected mobile learning application, KoTBaM and Learning Fakih. This study involves deaf children aged 7 – 12 years old

familiar with the mobile application. The survey is divided into two sections: i) demographic information and ii) 24 questions that the respondent must answer using a smileyometer. The survey included 38 deaf children from Malaysian Deaf School. The participating deaf children completed the questionnaires with the assistance of their teachers after using the mobile learning application in the classroom. Yet, various issues needed to be addressed in order to improve the deaf children's user experience. Special exercises should be developed for deaf children connected to their school syllabus to consolidate their knowledge and self-learn everywhere. Furthermore, games elements should be adapted so the deaf children are able to learn while playing.

Author 1: Normala Mohamad Author 2: Nor Laily Hashim

Keywords: User experience; UX testing; UX dimension; UX metrics; mobile learning application; deaf children; smileyometer

DOWNLOAD PDF

Paper 35: The Regularization Effect of Pre-activation Batch Normalization on Convolutional Neural Network Performance for Face Recognition System Paper

Abstract: Face recognition is of pronounced significance to real-world applications such as video surveillance systems, human computing interaction, and security systems. This biometric authenticating system encompasses rich real human face characteristics. As such, it has been one of the important research topics in computer vision. Face recognition systems based on deep learning approaches suffer from internal covariate shift problems that cause gradients to explode or gradient disappearance, which leads to improper network training. Improper network training causes network overfitting and computational load. This reduces recognition accuracy and slows down network speed. This paper proposes a modified pre-activation batch normalization convolutional neural network by adding a batch normalization layer after each convolutional layer within each of the four convolutional units of the proposed model. The performance of the proposed model is validated with a new dataset, AS-Darmaset, which is built out of two publicly available databases. This paper compared the convergence behavior of four different CNN models: the Pre-activation Batch Normalization CNN model, the Traditional CNN without Batch Normalization, the Post-Activation Batch Normalization CNN model, and the Sparse Batch Normalization CNN Architecture. The evaluation results show that the recognition performance of Pre-activation BN CNN has training and validation accuracies of 100.00% and 99.87%, the Post activation Batch normalization has 100.00% and 99.81%, and the traditional CNN without BN has 96.50% and 98.93%. The sparse batch normalization CNN has 96.25% and 97.60% success rate, respectively. The result shows that the Pre-activation BN CNN model is more effective than the other three deep learning models.

Author 1: Abu Sanusi Darma

Author 2: Fatma Susilawati Binti Mohamad

Keywords: Face recognition; pre-active batch normalization; convolutional neural network

DOWNLOAD PDF

Paper 36: Heuristics and Think-aloud Method for Evaluating the Usability of Game-based Language Learning

Abstract: Digital learning environments are increasingly becoming popular in recent years. The rising usage of cell phones has invited researchers to design and develop learning applications and games for mobile phones. Specifically, game-based language learning is being promoted by researchers in many parts of the world. “Language Learning Serious Game (LLSG)” is based on a theoretical model constructed by the researcher that supports children learning English as a second language in a cultural context. The usability of such games is evaluated based on well-defined heuristics and other standard methods. This research aims to appraise the usability of LLSG through heuristics and think-aloud approaches while involving all essential stakeholders, including language experts, students, teachers, and game developers. The researcher proposed the heuristics in a cultural context, whereas the think-aloud review is compiled from the rigorous discussion session involving these stakeholders to evaluate the LLSG. The findings obtained from the heuristics evaluation reveal that the usability of LLSG is acceptable. On the other hand, various interesting suggestions and reviews were gathered from the discussion between experts and students. This evaluation will further improve the future versions of the game.

Author 1: Kashif Ishaq Author 2: Fadhilah Rosdi

Author 3: Nor Azan Mat Zin Author 4: Adnan Abid

Keywords: Engagement; game-based; heuristic evaluation; language learning; motivation; think aloud; usability

DOWNLOAD PDF

Paper 37: Towards Measuring User Experience based on Software Requirements

Abstract: User Experience (UX) provides insights into the users’ product perceptions while using or intending to use an application. Software products are known for complexity and changeability, starting from requirements engineering until the product operation. Users often evaluate software UX based on a prototype; however, UX is semantically embedded in the software requirements, a crucial indicator for project success. The problem of current UX evaluation methods is their dependence on the actual involvement of users or experts, a time-consuming process. First, this paper builds a benchmark dataset of UX based on textual software requirements crowdsourcing several UX experts. Second, the paper develops a machine learning model to measure UX based on the dataset. This research describes the dataset characteristics and reports its statistical internal consistency and reliability. Results indicate a high Cronbach Alpha and a low root mean square error of the dataset. We conclude that the new benchmark dataset could be used to estimate UX instantly without the need for subjective UX evaluation. The dataset will serve as a foundation of UX features for machine learning models.

Author 1: Issa Atoum Author 2: Jameel Almalki

Author 3: Saeed Masoud Alshahrani

Author 4: Waleed Al Shehri

Keywords: User experience; benchmark dataset; requirements engineering; UX evaluation; software engineering

DOWNLOAD PDF

Paper 38: Machine Learning Driven Feature Sensitive Progressive Sampling Model for BigData Analytics

Abstract: BigData requires processing a huge data volume, which is an undeniable challenge for academia-industries. The classical sampling techniques are limited when addressing data-imbalance, large data-heterogeneity, multi-dimensionality etc. To alleviate it, in this paper a novel machine learning driven feature sensitive progressive sampling (ML-FSPS) that in conjunction with an improved feature selection and classification environment achieves more than 95.7% of accuracy, even with 10-14% of the original data size. The proposed ML-FSPS model was applied for IoT-device classification problem that possesses exceedingly high data-imbalance, multi-dimensionality and heterogeneity issues. Functionally, the FSPS-driven analytics model at first performed active period segmentation followed by multi-dimensional (descriptive) statistical feature extraction and Wilcoxon Rank Sum Test based feature selection. Subsequently, it executed K-Means clustering over a gigantically huge feature instances (16,00,000,000 network traces) Here, K-means algorithm clustered each feature samples into five distinct clusters. With initial sample size of 10%, FSPS model selected same amount of data elements (0.5-5% iteratively) from each cluster for each feature to perform multi-class classification using homogenous ensemble learning (HEL) model. Here HEL encompassed AdaBoost, Random Forest and Extended Tree ensemble algorithms as base classifiers. The simulation results affirmed that the proposed model achieves accuracy of almost 99% even with 10-16% of sample size.

Author 1: Nandita Bangera Author 2: Kayarvizhy N

Keywords: Feature sensitive progressive sampling; BigData analytics; machine learning; ensemble learning; rank sum test; IoT-device classification

DOWNLOAD PDF

Paper 39: Thermal-aware Dynamic Weighted Adaptive Routing Algorithm for 3D Network-on-Chip

Abstract: 3D Network-on-Chip NoC based systems have severe thermal problems due to the stacking of dies and disproportionate cooling efficiency of different layers. While adaptive routing can help with thermal issues, current routing algorithms are either thermally imbalanced or suffer from traffic congestion. In this work a novel thermal aware dynamic weighted adaptive routing algorithm has been proposed that takes traffic and temperature information into account and prevents packets being routed across congested and thermally aggravated areas. Dynamic weighted model will consider parameters related to congestion and thermal issues and provide a balanced suitable approach according to the current scenario at each node. The efficiency of the proposed algorithm is analyzed and evaluated with state-of-the-art thermal-aware routing algorithms using a simulation environment. Results obtained from the simulation shows that the proposed algorithm has performed better in terms of global average delay with 17-33 percent

improvement and better thermal profiling under various synthetic traffic conditions.

Author 1: Muhammad Kaleem Author 2: Ismail Fauzi Bin Isnin

Keywords: Routing algorithms; thermal-aware; dynamic weighted model; 3D Network-on-Chip

DOWNLOAD PDF

Paper 40: A New Back-off Algorithm with Priority Scheduling for MQTT Protocol and IoT Protocols

Abstract: The Internet of Things (IoT) protocols have encountered great challenges as the growth of technology has led to many limitations of the performance of the IoT protocols. Message Queuing Telemetry Transport protocol (MQTT) is one of the most dominant protocols in most fields of smart applications, so it has been chosen in this research to be a use case for implementing and evaluating a new proposed Back-off algorithm that is designed to eliminate suspicious and fake messages by calculating an initial frequent rate for each publisher connected to the MQTT broker. The proposed Back-off algorithm was designed to mitigate the traffic load of the uplink traffic by applying an exponential delay factor to suspicious publishers. Another priority scheduling algorithm was proposed to classify publishers as high priority or low priority depending on the new calculated frequent rate. The two algorithms were implemented on the Mosquitto broker and evaluated using a simulation environment by measuring specified performance metrics. The simulated results proved that the Back-off algorithm eliminated network load and introduced an acceptable range of CPU and RAM consumption. The results also concluded that the priority classification algorithm managed to reduce the latency of high-priority publishers.

Author 1: Marwa O Al Enany Author 2: Hany M. Harb

Author 3: Gamal Attiya

Keywords: Back-off algorithm; priority scheduling; MQTT protocol; average transmission frequency rate; IoT protocols

DOWNLOAD PDF

Paper 41: Examining User Experience of Moodle e-Learning System

Abstract: This research investigates the user experience (UX) of the Moodle e-learning system employed at one university in Malaysia from students' perspectives. Comprehensive user experience (UX) criteria were suggested, which was adopted from two reliable criteria, to evaluate the user experience (UX) of the e-learning system. The suggested comprehensive user experience (UX) criteria consist of 8 categories and 29 corresponding sub-categories; these can be used to evaluate teaching and learning, usability, and hedonic aspects of an e-learning system. Semi-structured interviews and questionnaires were employed based on the suggested user experience (UX) criteria to collect qualitative and quantitative data regarding users' experience (UX) of the tested e-learning system. The results showed that the e-learning system had positive user experience (UX) in general from the students'

perspectives. The results also showed that the students were satisfied with most of the metrics related to teaching and learning, usability and hedonic. However, the students identified some challenges they faced while interacting with the e-learning system which could be improved in order to improve their user experience (UX) and gain more benefits from a good user experience (UX) e-learning system.

Author 1: Layla Hasan

Keywords: User Experience (UX); e-learning system; Moodle; usability; learning management system

DOWNLOAD PDF

Paper 42: Query Expansion based on Word Embeddings and Ontologies for Efficient Information Retrieval

Abstract: Information retrieval has been an ever-going process for end users to fetch relevant data at one go. The problem intensifies more with unstructured data in a semantic web environment. It is also a promising area for researchers to dive in and refine it from time to time. Expanding the user query and reformulating it is one probable solution to increase the efficiency of the information retrieval system. In this paper we propose “WeOnto”, a novel two-level query expansion algorithm that utilizes the combination of web ontologies and word embeddings for similarity calculation. In the first level, the Real estate Ontology (REO) is created using Protégé and Sparql queries are passed to retrieve probable semantic words from the given ontology for each inputted user query. The first level gave significant results and improved the information retrieval by 18%. The second level of algorithm uses word embedding enhanced with the domain knowledge that helps to retrieve similar meaningful words based on cosine similarity for the same user query. Word embeddings are implemented using Word2Vec method that follows two architectures namely CBOW or Skip Gram. Most similar semantic words are retrieved using the CBOW word embeddings method in the proposed algorithm and concatenated with the semantic keywords generated from the real estate ontology to form a powerful reformulated query that gives promising relevant results. Finally, two topmost words as per their similarity index are taken to reformulate the original user query. Experimental results depict that proposed algorithm has given distinct results and has showcased significant improvement of 93% over the initial user query.

Author 1: Namrata Rastogi Author 2: Parul Verma

Author 3: Pankaj Kumar

Keywords: CBOW; Information retrieval; ontology; query reformulation; semantic web; skip gram; word embeddings; word2vec

DOWNLOAD PDF

Paper 43: A Novel Integrated Scheme for Detection and Mitigation of Route Diversion Attack in MANET

Abstract: With the involvement of Mobile Adhoc Network (MANET) in many upcoming technologies and applications, there is an increasing concern about secure data transmission. Until the last decade, various solutions have evolved

to circumvent this threat; however, the security issue is still a more significant threat. The problems studied during the review are usage of Complex Cryptographic Usage, Less Energy Efficient, Fewer studies towards Route Diversion Attack, and Less Emphasis towards Securing Beacon. An analytical method has been used to study these problems. This paper introduces a novel scheme that carries out dual operation viz. i) assessing the link legitimacy for detection of route diversion attack, and ii) cost-effective countermeasures for the same attack. The key findings of proposed study is token generation process when associated with link legitimacy offers more routing security from various ranges of threats. The broader implication of this finding is that proposed system when characterized by lightweight encryption operation, it is capable of excelling better balance between data transmission and security performance unlike existing security solutions in MANET.

Author 1: H C Ramaprasad Author 2: S. C. Lingareddy

Keywords: Mobile Adhoc network; route diversion attack; routing attack; link legitimacy; encryption

DOWNLOAD PDF

Paper 44: Multi-level Hierarchical Controller Assisted Task Scheduling and Resource Allocation in Large Cloud Infrastructures

Abstract: The high-pace emergence in Cloud Computing technologies demands and alarmed academia-industries to attain Quality-of-Service (QoS) oriented solutions to ensure optimal network performance in terms of Service Level Agreement (SLA) provision as well as Energy-Efficiency. Majority of the at-hand solutions employ Virtual Machine Migration to perform dynamic resource allocation which fails in addressing the key problem of SLA-sensitive scheduling where it demands timely and reliable task-migration solution. Undeniably, VM consolidation may help achieve energy-efficiency along with dynamic resource allocation where the classical heuristic methods which are often criticized for its local minima and premature convergence doesn't guarantee the optimality of the solution, especially over large cloud infrastructures. Considering these key problems as motivation, in this paper a highly robust and improved meta-heuristic model based on Ant Colony System is developed to achieve Task Scheduling and Resource Allocation. CloudSim based simulation over different PlanetLab cloud traces exhibited superior performance by the proposed task-scheduling model in terms of negligible SLA violence, minimum downtime, minimum energy-consumption and higher number of migrations over other heuristic variants, which make it suitable towards realistic Cloud Computing purposes.

Author 1: Jyothi S Author 2: B S Shylaja

Keywords: Task-scheduling; VM-migration; improved ant colony system; SLA assurance; energy-efficient consolidation

DOWNLOAD PDF

Paper 45: A Review of a Biomimicry Swimming Robot using Smart Actuator

Abstract: Biomimicry-based robotic mobility is a newer subgenre of bio-inspired design and it's all about applying natural concepts to the development of real-world engineered systems. Previously, researchers used actuators such as motors, pumps, and intelligent materials or intelligent actuators to build many biomimicry robots. Due to the field's growing interest, this study will examine the performance of several biomimicry robots that have been built based on their different design, the type of material the robot utilizes, and the type of propulsion for the robot to swim while providing huge thrust. Robots must not only design such an animal, but its maneuverability and control tactics must also be tied to wildlife to provide the finest impersonation of biological life. Fish propulsion can be separated into two categories which are body and/or caudal fins (BCF) and median and/or paired fins (MPF). The old propeller system in underwater robot usually uses motor and pump. Many researchers have begun developing smart materials as drivers in recent years that can be grouped into four categories: shape memory alloy SMA, ionic polymer metal composite IPMC, lead zirconate titanate (PZT) and pneumatic soft actuator as replacement for pump or motor. Varied materials produce different result and can be applied for different propulsion modes. Future researchers working on biomimetic fish robots will be guided by the findings of this study.

Author 1: Muhammad Shafique Ashroff Md Nor

Author 2: Mohd Aliff Author 3: Nor Samsiah

Keywords: Biomimicry; fish propulsion; biological life; smart actuator

DOWNLOAD PDF

Paper 46: Fuel Consumption Prediction Model using Machine Learning

Abstract: In the paper, we are enhancing the accuracy of the fuel consumption prediction model with Machine Learning to minimize Fuel Consumption. This will lead to an economic improvement for the business and satisfy the domain needs. We propose a machine learning model to predict vehicle fuel consumption. The proposed model is based on the Support Vector Machine algorithm. The Fuel Consumption estimation is given as a function of Mass Air Flow, Vehicle Speed, Revolutions Per Minute, and Throttle Position Sensor features. The proposed model is applied and tested on a vehicle’s On-Board Diagnostics Dataset. The observations were conducted on 18 features. Results achieved a higher accuracy with an R-Squared metric value of 0.97 than other related work using the same Support Vector Machine regression algorithm. We concluded that the Support Vector Machine has a great effect when used for fuel consumption prediction purposes. Our model can compete with other Machine Learning algorithms for the same purpose which will help manufacturers find more choices for successful Fuel Consumption Prediction models.

Author 1: Mohamed A. HAMED

Author 2: Mohammed H.Khafagy Author 3: Rasha M.Badry

Keywords: Fuel consumption; machine learning; support vector machine; feature weight; feature selection; on-board diagnostic

DOWNLOAD PDF

Paper 47: Scalable and Reactive Multi Micro-Agents System Middleware for Massively Distributed Systems

Abstract: IT transformation has revolutionized the business landscape and changed most of organizations business model into digital and innovation driven firms. To fully take advantage of this digitalization and the exponential growth of data, organizations need to rely on resilient, scalable, extremely connected, highly available & very performant systems. To meet this need, this paper presents a model of middleware for multi micro-agents system based on reactive programming and designed for massively distributed systems and High-Performance Computing, especially to face big data challenges. This middleware is based on multi-agents systems (MAS) which are known as a reliable solution for High Performance Computing. This proposal framework is built on abstraction and modularity principles through a multi-layered architecture. The design choices aim to ensure cooperation between heterogeneous distributed systems by decoupling the communication model and the cognitive pattern of micro agents. To ensure high scalability and to overcome networks latency, the proposal architecture uses distribution model of data & computing, that allows an adaptation of the grid size as needed. The resilience problem is addressed by adopting the same mechanism as Hazelcast middleware, thanks to his peer-to-peer architecture with no single point of failure.

Author 1: EZZRHARI Fatima Ezzahra

Author 2: EL ABID AMRANI Noureddine

Author 3: YOUSSFI Mohamed Author 4: BOUATTANE Omar

Keywords: Massively distributed system; multi agent system (MAS); high performance computing; reactive programming; hazelcast

DOWNLOAD PDF

Paper 48: Expert System in Enhancing Efficiency in Basic Educational Management using Data Mining Techniques

Abstract: The importance of basic education is well noted in every country. Proper planning and utilization of resources at the basic level, helps in leveraging the success of education at all other levels of education in a country. Ghana is noted to be the country that spends higher in education than its West African counterparts. In Ghana, attempts made to plan by predicting and projecting expenditure as well as the available resources to manage basic education are not accurate enough to address the challenges of education in the country. With the issue of COVID 19 pandemic, more expenditure is realized in managing educational institutions as more resources are needed in observing the protocols to curtail the pandemic. This throws a serious challenge to the effective and efficient utilization of the limited resources in the country. In this paper, the data from the Ministry of Education is analysed using data mining techniques. This has helped to identify the inaccuracies in the data. Inaccurate population projection affects the Key Performance Indicators (KPIs) in education because population is a common denominator for educational indicators. A proposed expert system is to be developed to assist in managing the situation.

Author 1: Fuseini Inusah Author 2: Yaw Marfo Missah

Author 3: Najim Ussiph Author 4: Frimpong Twum

Keywords: Basic education; data mining; educational management; expert system; population

DOWNLOAD PDF

Paper 49: Machine Learning for Predicting Employee Attrition

Abstract: Employee attrition has become a focus of researchers and human resources because of the effects of poor performance on organizations regardless of geography, industry, or size. In this context, the use of machine learning classification models to predict whether an employee is likely to quit could greatly increase the human resource department's ability to intervene on time and possibly provide a remedy to the situation to prevent attrition. This study is conducted with an objective to compare the performance machine learning techniques, namely, Decision Tree (DT) classifier, Support Vector Machines (SVM) classifier, and Artificial Neural Networks (ANN) classifier, and select the best model. These machine learning techniques are compared using the IBM Human Resource Analytic Employee Attrition and Performance dataset. Preprocessing steps for the dataset used in this comparative study include data exploration, data visualization, data cleaning and reduction, data transformation, discretization, and feature selection. In this study, parameter tuning and regularization techniques to overcome overfitting issues are applied for optimization purposes. The comparative study conducted on the three classifiers found that the optimized SVM model stood as the best model that can be used to predict employee attrition with the highest accuracy percentage of 88.87% as compared to the other classification models experimented with, followed by ANN and DT.

Author 1: Norsuhada Mansor Author 2: Nor Samsiah Sani

Author 3: Mohd Aliff

Keywords: Artificial neural networks; decision tree; employee attrition; machine learning; support vector machines

DOWNLOAD PDF

Paper 50: Finding Good Binary Linear Block Codes based on Hadamard Matrix and Existing Popular Codes

Abstract: Because of their algebraic structure and simple hardware implementation, linear codes as class of error-correcting codes, are used in a multitude of situations such as Compact disk, backland bar code, satellite and wireless communication, storage systems, ISBN numbers and so more. Nevertheless, the design of linear codes with high minimum Hamming distance to a given dimension and length of the code, remains an open challenge in coding theory. In this work, we propose a code construction method for constructing good binary linear codes from popular ones, while using the Hadamard matrix. The proposed method takes advantage of the MacWilliams identity for computing the weight distribution, to overcome the problem of computing the minimum Hamming distance for larger dimensions.

Author 1: Driss Khebbou Author 2: Reda Benkhouya

Author 3: Idriss Chana Author 4: Hussain Ben-azza

Keywords: Binary linear codes; code construction; minimum hamming distance; error-correcting codes; weight distribution; coding theory; hadamard matrix

DOWNLOAD PDF

Paper 51: EFPT-OIDS: Evaluation Framework for a Pre-processing Techniques of Automatic Optho-Imaging Diagnosis and Detection System

Abstract: The modalities of FUNDUS images and the availability of public domain data sets provides a starting point in designing an ecosystem for developing an automatic detection of degenerative early-stage Glaucoma and Diabetic Retinopathy, and other eye-related diseases. The existing techniques for these operations lack flexibility and robustness in their design implementation and are limited to only certain preprocessing requirements. However, the existing methods are useful but provide lower performance when the FUNDUS image quality degrades due to misalignment of lens opening in camera and poor functioning of visual sensors. This paper presents a unified framework that mechanizes different preprocessing techniques to benefit the Optho-imaging diagnosis and disease detection process. The proposed framework facilitates on-demand data treatment operations that include image interpolation, brightness adjustment, illumination correction, and noise reduction. The proposed techniques for FUNDUS image enhancement provide better PSNR and SSIM-performance metrics for image quality than existing popular image enhancement techniques when tested on two standard publicly available datasets. The contribution of the proposed framework is that it offers flexible and effective mechanisms that meet dynamic preprocessing operations on an on-demand basis to prepare better data representation for building machine learning models. The framework can also be used in real-time for eye disease diagnosis by an ophthalmologist.

Author 1: Sobia Naz Author 2: Radha Krishna Rao K. A

Author 3: Shreekanth T

Keywords: Pre processing; FUNDUS image; glaucoma; diabetic retinopathy; interpolation; image enhancement

DOWNLOAD PDF

Paper 52: A Delay-tolerant MAC Protocol for Emergency Care in WBAN Considering Preemptive and Non-preemptive Methods

Abstract: For facilitating pilgrims with no delay, quick and real-time emergency medical services at ritual sites a delay-tolerant Medium Access Control (MAC) protocol for the IEEE 802.15.6 standard based Wireless Body Area Networks (WBANs) has been proposed. Since MAC protocol is application-specific hence any particular MAC technique may not be appropriate for diverse applications. In this research work, we consider dealing with medical emergency traffics which is random, independent of each other and can be generated at any time. Moreover, emergency traffics must be transmitted ahead of normal medical data or emergency traffic with a lower severity level; because any delay in emergency data transmission may endanger patients'

life. The proposed MAC protocol is compared with both preemptive and non-preemptive methods. Where, a modified MAC superframe (SF) structure, minimum backoff period and minimum Contention Window (CWmin) for quick data access to the IEEE 802.15.6 standard based EAP channel are also considered. The proposed delay-tolerant MAC protocol has been experimented with and simulated by the Castalia simulator which is based on the OMNeT++ platform. The experimental results show that data transmission using the preemptive method works faster with reduced delay than that of the non-preemptive method. Furthermore, the delay metric of the proposed delay-tolerant MAC protocol is analyzed, calculated and compared with the current Traffic-aware TA-MAC protocol. Results demonstrate that delay is relatively low during emergency data transmission using the proposed MAC in WBANs environment.

Author 1: Shah Murtaza Rashid Al Masud

Author 2: Aloke Kumar Saha

Keywords: WBAN; MAC; preemptive; non-preemptive; delay; emergency traffic

DOWNLOAD PDF

Paper 53: Transformer based Contextual Model for Sentiment Analysis of Customer Reviews: A Fine-tuned BERT

Abstract: The Bidirectional Encoder Representations from Transformers (BERT) is a state-of-the-art language model used for multiple natural language processing tasks and sequential modeling applications. The accuracy of predictions from context-based sentiment and analysis of customer review data from various social media platforms are challenging and time-consuming tasks due to the high volumes of unstructured data. In recent years, more research has been conducted based on the recurrent neural network algorithm, Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM) as well as hybrid, neutral, and traditional text classification algorithms. This paper presents our experimental research work to overcome these known challenges of the sentiment analysis models, such as its performance, accuracy, and context-based predictions. We've proposed a fine-tuned BERT model to predict customer sentiments through the utilization of customer reviews from Twitter, IMDB Movie Reviews, Yelp, Amazon. In addition, we compared the results of the proposed model with our custom Linear Support Vector Machine (LSVM), fastText, BiLSTM and hybrid fastText-BiLSTM models, as well as presented a comparative analysis dashboard report. This experiment result shows that the proposed model performs better than other models with respect to various performance measures.

Author 1: Ashok Kumar Durairaj Author 2: Anandan Chinnalagu

Keywords: Transformers model; BERT; sequential model; deep learning; RNN; LSVM; LSTM; BiLSTM; fastText

DOWNLOAD PDF

Paper 54: A New Energy-efficient Multi-hop Routing Protocol for Heterogeneous Wireless Sensor Networks

Abstract: Energy use of sensor nodes efficiently and extending the lifetime of heterogeneous wireless sensor networks (HWSNs) is a main goal of HWSNs routing optimization methods, and therefore building an energy-efficient routing protocol becomes critical for HWSN performance improvement. They present an energy-efficient routing protocol based on the grey wolf optimizer (GWO) and the Tabu search algorithm (TSA) in this paper. Proposed routing system with primary objectives include clustering and the selection of cluster heads (CH) utilizing GWO with a fitness function based on the residual energy of sensor nodes and the average distance between CH and sink nodes base station (BS) due to the mobility of sensors, the quality of service (QoS) parameters such as reliability and energy consumption can be improved by discovering multiple optimized paths for data transmission from CH to BS and by TSA selecting the optimal route from CH to BS based on the forwarding of reliable route packets (FRRPs). The experimental results indicate that the proposed grey wolf optimizer with tabu Search Algorithm (GWO-TSA) can reduce HWSNs energy consumption by 10% and 20%, increase lifetime by 13% and 18%, and finally increase throughput by 6% and 14% when compared to the genetic algorithm with tabu search algorithm (GA-TSA) and grey wolf optimizer with crow search algorithm (GWO-CSA). When compared to GA-TSA & GWO-CSA, simulation reveals that the proposed GWO-TSA protocol improves HWSNs performance by minimizing energy consumption, maximizing network lifetime, and boosting network throughput.

Author 1: Rowayda A. Sadek Author 2: Doha M. Abd-alazeem

Author 3: Mohamed M. Abbassy

Keywords: Heterogeneous wireless sensor networks (HWSNs); forwarding of reliable route packets (FRRPs); grey wolf optimizer (GWO); routing optimization; tabu search algorithm (TSA); quality of service (QoS)

DOWNLOAD PDF

Paper 55: Improved GRASP Technique based Resource Allocation in the Cloud

Abstract: In the era of cloud computing, everyone is somehow using cloud resources. However, the resources are limited in the Cloud. Cloud vendors look for enhanced returns on investments. Promising return on investment is possible only when the cloud resources are scheduled efficiently to execute jobs within the stipulated time. However, brute force methods require exponential time to produce a schedule. Heuristic and meta-heuristic algorithms have been proposed in the literature to allocate resources to the jobs. These algorithms still suffer from slow convergence. To overcome this problem, researchers clubbed various heuristics and meta-heuristic to form a new hybrid algorithm. With the same motive, this paper explores the limitations of greedy random adaptive search and shows that learning through a fixed set search enhances efficiency. Based on the results, it can be concluded that the proposed algorithm is on par with existing hybrid meta-heuristic algorithms.

Author 1: Madhukar E Author 2: Ragunathan T

Keywords: Cloud computing; task scheduling; meta-heuristics; fixed set search; GRASP; resource allocation

DOWNLOAD PDF

Paper 56: Ada-IDS: AdaBoost Intrusion Detection System for ICMPv6 based Attacks in Internet of Things

Abstract: The magical buzzword Internet of Things (IoT) connects any objects which are diverse in nature. The restricted capacity, heterogeneity and large scale implementation of the IoT technology tend to have lot of security threats to the IoT networks. RPL is the routing protocol for the constraint devices like IoT nodes. ICMPv6 protocol plays a major role in constructing the tree-like topology called DODAG. It is vulnerable to several security attacks. Version Number Attack, DIS flooding attack and DAO attack are the ICMPv6 based attacks discussed in this paper. The network traffic is collected from the simulated environment in the normal and attacker settings. An AdaBoost ensemble model termed Ada-IDS is developed in this research to detect these three ICMPv6 based security attacks in RPL based Internet of Things. The proposed model detects the attacks with 99.6% accuracy and there is no false alarm rate. The Ada-IDS ensemble model is deployed in the Border Router of the IoT network to safeguard the IoT nodes and network.

Author 1: A. Arul Anitha Author 2: L. Arockiam

Keywords: IoT; ICMPv6; version number attack; DIS attack; DAO attack; Ada-IDS

DOWNLOAD PDF

Paper 57: Analytical Framework for Binarized Response for Enhancing Knowledge Delivery System

Abstract: The student feedback offer effective insight into their experience of knowledge transfer, routinely collected in academic institutions. However, the existing research literature lacks reporting whether the comments in education system are helpful or non-useful. Most of the existing works are limited to sentiment polarity computation only, and teacher evaluation is carried out without considering the aspects of the teaching. This study analyzes student comments and classifies comments as applicable and non-useful for the teacher scoring system. In the proposed research, the data considered is the student feedback collected from the teacher rating website. The study performed phase-by-phase text data modeling. First, exploratory analysis is carried out on the student feedback dataset to understand text data characteristics and features. Based on the exploratory analysis, appropriate steps are determined to perform preprocessing operations for data cleaning. Using natural language processing context, the study only focuses on removing stop words and common words that belong to both useful and non-useful contexts. BoW model is considered for features extraction, and two probabilistic supervised machine learning models are used for comment classification. The study outcome exhibits that Gaussian Naïve Bayes outperforms Multinominal Naïve Bayes in accuracy, precision, recall rate, and F1-score.

Author 1: Chethan G S Author 2: Vinay S

Keywords: Education; knowledge transfer; machine learning; natural language processing; student feedback

DOWNLOAD PDF

Paper 58: A Linguistic Analysis Metric in Detecting Ransomware Cyber-attacks

Abstract: Originating and striking from anywhere, cyber-attacks have become ever more sophisticated in our modern society and users are forced to adopt increasingly good and vigilant practices to protect from them. Among these, ransomware remains a major cyber-attack whose major threat to end users (disrupted operations, restricted files, scrambled sensitive data, financial demands, etc.) does not particularly lie in number but in severity. In this study we explore the possibility of real-time detection of ransomware source through a linguistic analysis that examines machine translation relative to the Levenshtein Distance and may thereby provide important indications as to attacker's language of origin. Specifically, the aim of our research is to advance a metric to assist in determining whether an external ransom text is an indicator of either a human- or a machine-generated cyber-attack. Our proposed method works its argument on a set of Eastern European languages but is applicable to a large(r) range of languages and/or probabilistic patterns, being characterized by usage of limited resources and scalability properties.

Author 1: Diana Florea Author 2: Wayne Patterson

Keywords: Cyber-security; cyber-attacks; machine translation; language; Levenshtein distance

DOWNLOAD PDF

Paper 59: Design and Implementation of Dynamic Packet Scheduling with Waiting Time Aware: DPSW2A

Abstract: One of the principal goals of 5G is to enhance performance in connection with speed and delay curtailment. To accomplish this, IETF has proposed Multipath TCP to utilize the accessible interface for communication. The demand for mobile communication is escalating day by day. The predominant communication option for people is mobile. For giving better service for users', nodes are fitted out with multiple interfaces. Multiple interfaces are as well one of the benefits of 5G. Multi path protocols are used to load balancing and resilience to failure. When communicating with asymmetric interfaces, latency is an imperative factor. To attain low latency is hard when asymmetric interfaces are used. When communication happens using multiple interfaces, the scheduler plays a central role since it decides which interface needs to be used for the packet. In this article we spotlight on scheduling algorithms, how this schedule will play a vital role to transfer data to receiver nodes with low latency. In this paper, we emphasize on the Scheduler named DPSWWA with the objective of minimizing delay and effective usage of interfaces.

Author 1: K Raghavendra Rao Author 2: B N Jagadesh

Keywords: MPTCP; scheduler; latency; delay; transport protocol; waiting time

DOWNLOAD PDF

Paper 60: Evaluating Domain Knowledge and Time Series Features for Automated Detection of Schizophrenia from EEG Signals

Abstract: Over the recent years, Schizophrenia has become a serious mental disorder that is affecting almost 21 million people globally. There are different symptoms to recognize schizophrenia from healthy people. It can affect the thinking pattern of the brain. Delusions, hallucinations, and disorganized speech are the common symptoms of Schizophrenia. In this study, we have used electroencephalography (EEG) signals to analyze and diagnose Schizophrenia using machine learning algorithms and found that temporal features performed well as compared to statistical features. EEG signals are the best way to analyze this disorder as they are intimately linked with human thinking patterns and provide information about brain activities. The present work proposes a Machine Learning (ML) model based on Logistic Regression (LR) along with two feature extraction libraries Time Series Feature Extraction Library (TSFEL) and MNE Python toolkit to diagnose Schizophrenia from EEG signals. The results are analyzed based on 5 different sampling techniques. The dataset was cross-validated using leave one subject out cross-validation (LOSOCV) using Scikit learn and achieve greater accuracy, sensitivity, specificity, macro average recall, and macro f1 score on temporal features respectively.

Author 1: Saqib Hussain Author 2: Nasrullah Pirzada

Author 3: Erum Saba Author 4: Muhammad Aamir Panhwar

Author 5: Tanveer Ahmed

Keywords: Artificial intelligence (AI); Logistic Regression (LR); smote-class weight (S-CW); borderline smote-class weight (BS-CW); electroencephalography; Time Series Feature Extraction Library (TSFEL)

DOWNLOAD PDF

Paper 61: Long Term Solar Power Generation Prediction using Adaboost as a Hybrid of Linear and Non-linear Machine Learning Model

Abstract: The usage of renewable energy sources has increased manifolds in terms of electric utilities. From other non-conventional sources, solar energy has been seen as a promising and convenient source used around the globe. In terms of meeting the daily requirements, solar energy has huge potential to fulfil the world’s demand. However, firstly the characteristic of solar energy is very unpredictable and intermittent due to variation of weather. Secondly, the optimization and the planning of smart grid effect the operation of PV system. Thus, prediction on the long horizon is needed to address this problem. Nevertheless, long term forecasting of solar power generation is deliberated as a challenging problem. Therefore, this paper proposes a 10 day ahead solar power forecasting using combination of linear and non-linear machine learning models. At first, the outputs are generated from Recurrent Neural Network (RNN), Support Vector Machine (SVM) and Autoregressive with exogenous variable (ARX). Later on, these three outputs are combined and are made as a strong classifier with the Adaptive boost (Adaboost) algorithm. The simulations were conducted on the data obtained from real PV plant. By the experimental results and discussion, it was endogenously concluded that the combination of all techniques with the Adaboost have increased the performances and showing the high accuracy as compare to the individual machine learning models. The hybrid Adaboost shows %MAPE 8.88, which proven high accuracy. While on the other hand, for the individual technique, RNN shows 10.88, SVM reveals 11.78 and ARX gives 13.00 of percentage MAPE. The

improvement proves that combination of techniques performs better than individual models and proclaims the high accuracy.

Author 1: Sana Mohsin Babbar Author 2: Chee Yong Lau

Author 3: Ka Fei Thang

Keywords: Recurrent neural network (RNN); support vector machine (SVM); autoregressive with exogenous variable (ARX); adaptive boosting (Adaboost) and photovoltaic system (PV)

DOWNLOAD PDF

Paper 62: HORAM: Hybrid Oblivious Random Access Memory Scheme for Secure Path Hiding in Distributed Environment

Abstract: Now-a-days in most of the sectors digitization has taken place to store data and process it easily with enhanced techniques. Online transactions produce very huge data daily in various sectors like health care, military, government office. To store huge data many firms, take the help of third-party organizations and store data on machines provided by them which creates new security issues. While performing operations on the data or accessing data metadata leakage may happen due to untrustworthy systems. This paper proposed hybrid oblivious random-access memory (HORAM) offers users to access their data from untrusted storage devices without sharing any information about their access patterns or techniques. Here random data block shuffling approach is used which helps in hiding storage policies about the user data blocks placement and preserving privacy of data. HORAM techniques perform pull-push operations on data in a parallel manner which in turn minimizes network overhead and reduces the execution time of operation. An extensive experimental analysis of the proposed system produces better results than weak and strong Federated oblivious random access memory (FEDORAM) respectively. The method is faster than weak FEDORAM and strong FEDORAM as it takes 0.96 seconds for communication with 5 servers whereas weak and strong FEDORAM takes 1.5 and 2 seconds respectively for reading and writing data.

Author 1: Snehalata Funde Author 2: Gandharba Swain

Keywords: HORAM; metadata; data blocks; privacy; block shuffling

DOWNLOAD PDF

Paper 63: Blockchain-based Secure Data Transmission for UAV Swarm using Modified Particle Swarm Optimization Path Planning Algorithm

Abstract: With the rapid development of unmanned aerial vehicles assisted applications enabled with a communication system, they are open to various malicious attacks. As a new form of flying things, they can access the network for better communication via the aerial base station. Most of the Unmanned aerial vehicles assisted flying objects' optimal path selection schemes does not consider the path deviation. In path deviation attacks, secure data transmission are not addressed in existing works. The secure communication process between Unmanned aerial vehicles and base station are exploited through security-based attacks. Moreover, path loss issue leads to multicast packet

loss and unsecured broadcast. The existing network architecture setup does not fulfill the secure data communication and privacy issues. In this paper, Blockchain is utilized to investigate the secure communication between Unmanned aerial vehicles to Wireless Unmanned aerial vehicles Base stations. Since the destination information is dynamic under an uncertain environment, it will cause a delay in data communication. Unmanned aerial vehicles are more vulnerable to security attacks. The proposed blockchain-based architecture supports secure data communication in Unmanned aerial vehicles uncertain environments. To improve network security, this paper designs a modified particle swarm optimization method for better path selection. Through these experimental results, a blockchain-based data communication scheme is outer performed concerning network security.

Author 1: M. Kayalvizhi Author 2: S. Ramamoorthy

Keywords: Unmanned aerial vehicles; path planning; swarm optimization; denial of service attack; blockchain; security and privacy; data communication

DOWNLOAD PDF

Paper 64: Resource Allocation in Spectrum Deployment for Cognitive Third-party Users

Abstract: Spectrum scarcity is a major challenge in wireless communication for next generation applications. The spectrum sharing and utilization of other user available spectrum is an optimal solution, which has outcome with a new mode of communication called cognitive network. Cognitive devices are capable of requesting and sharing free spectrum among each other in a communication range. The spectrums are shared among primary and secondary user (PU, SU). To extend the range of spectrum utilization, users from other clusters are requested in sharing of spectrum which is called as third party user (TSU). In detection of free spectrum a jamming based approach is proposed in recent work. Jamming approach generates a periodic jamming signal for TSU in sensing of free spectrum. The repetitive requesting of spectrum availability result in large jamming probability for engaged TSU users resulting in large delay. In minimizing the delay observed, in this paper, a new monitored jamming approach is proposed. Proposed Monitored jamming approach is develop in recent to the engagement of each TSU for communication and governing the jamming signal based on the spectrum engagement. The proposed approach minimizes the delay due to jamming signaling in existing system. The Experimental results obtained shows an enhancement to the system throughput, fairness factor and minimizes the delay metric for proposed approach.

Author 1: Arikatla Jaya Lakshmi Author 2: G. N. Swamy

Author 3: M. N. Giri Prasad

Keywords: Slot monitoring; spectrum sensing; primary user (PU); secondary user (SU); third-party spectrum utilization (TSU)

DOWNLOAD PDF

Paper 65: The Use of the Relational Concept in the Arabic Morphological Analysis

Abstract: The Arabic language differs from other natural languages in its structures and compositions. In this article we have developed an Arabic morphological analyzer. For this, we have used the relational concept in the database to build our Arabic morphological analyzer. This analyzer uses a set of tables which are linked together by relationships. These relations model certain numbers of compatibility rules between different affixes. Our morphological analysis have been trained and tested on the same databases. The tests of our new approach have given good results and the numbers obtained are very close to those of existing analyzer.

Author 1: Said Iazzi Author 2: Abderrazak Iazzi

Author 3: Saida Laaroussi Author 4: Abdellah Yousfi

Keywords: Arabic language rules; morphology; morphological analyzers; database; relational concept

DOWNLOAD PDF

Paper 66: Development of Wearable System to Help Preventing the Spread of Covid-19 in Public Indoor Area

Abstract: Smart wearables as a part of the Internet of Things nowadays gaining confidence in our daily lives because of its accessibility and simplicity. Today, with the outbreak of Coronavirus around the world, a smart wearable device can become another solution to help slowing the virus spreading by ensuring public health and social measures. In this paper, a system consisting of distance detector and touchless door access is proposed to help the personal, physical and social distancing measures practice in a public indoor area. A BLE positioning method based on RSSI localization is used to ensure the physical distancing around the user. WPA2 and MAC address-based authentication for the touchless door access is used to restrict and trace the visitor of the indoor area. The system is implemented in ESP microcontroller. A proof of concept is conducted to see if the functionality of the system already satisfied the public health and social measures practice. The results show that only registered devices can give a signal to open the door and the device can guarantee the physical distance around the user with 4.51% error in indoor area.

Author 1: Annisa Istiqomah Arrahmah

Author 2: Surya Ramadhan

Keywords: Wearable device; healthcare system; COVID-19; door-lock system; radio signal strength indicator (RSSI)

DOWNLOAD PDF

Paper 67: A Comparative Study of Segmentation Method for Computer-aided Diagnosis (CAD) Leukemia AML Subtype M0, M1, and M2

Abstract: A computer-based diagnosis model for Acute Myelogenous Leukemia (AML) is carried out using white blood cell image processing. The stages in computer-aided diagnosis (CAD) are included pre-processing, segmentation, feature extraction, and classification. The segmentation method

has many approaches, namely, clustering, region growing, and thresholding. The number of approaches that can be used requires proper selection because it will have an impact on CAD performance. This study aims to conduct a comparative study of the performance of the WBC segmentation method on the AML M0, M1, and M2 subtype leukemia CAD system. The segmentation algorithm used is k-means, fuzzy c-means, SOM, watershed, chan vese (active contour), otsu thresholding, and histogram. The feature extraction method uses GLCM, while the classification algorithms tested are SVM, Random-forest, decision tree, naive Bayesian, and k-NN. The test results show that the histogram segmentation method is able to provide the best average performance when using SVM, namely 90.3% accuracy, 85.9% sensitivity, and 92.7% specificity.

Author 1: Wiharto Author 2: Wisnu Widiarto

Author 3: Esti Suryani Author 4: Nurmajid Hidayatullah

Keywords: Acute myelogenous leukemia; leukemia; segmentation; feature extraction; classification

DOWNLOAD PDF

Paper 68: EC-Elastic an Explicit Congestion Control Mechanism for Named Data Networking

Abstract: In recent years, Named Data Networking (NDN) has attracted researchers' attention as a new internet architecture. NDN changes the internet communication paradigm from a host-to-host IP model to a name-based model. Thus, NDN permits the retrieval of requested content by name, from different sources and via multiple paths, and the use of caching in intermediate routers. These features transform the transport control model from sender to receiver and make traditional end-to-end congestion control mechanisms incompatible with NDN architecture. To deal with this problem, a reliable congestion control mechanism becomes necessary for a successful deployment of NDN. This paper presents a new hybrid congestion control mechanism for NDN, EC-Elastic (Explicit Congestion Elastic), which adopts the basic concept of Elastic-TCP to control the sending rates of the interest packets at the consumer nodes. In the intermediate nodes, a queue has been associated with the Controlled Delay-Active Queue Management CoDel-AQM to measure the packet sojourn time and notify the consumer to decrease its interest packet sending rate when it receives an explicit congestion signal. EC-Elastic was implemented in ndnSIM and evaluated with Agile-SD, CUBIC, and STCP in different scenarios. Simulation results show that EC-Elastic provides a significant improvement in bandwidth utilization while maintaining lower delay and packet loss rates.

Author 1: Asmaa EL-BAKKOUCHI

Author 2: Mohammed EL GHAZI Author 3: Anas BOUAYAD

Author 4: Mohammed FATTAH

Author 5: Moulhime EL BEKKALI

Keywords: NDN; named data networking; congestion control; explicit congestion control; TCP-elastic

DOWNLOAD PDF

Paper 69: 1D-CNN based Model for Classification and Analysis of Network Attacks

Abstract: With the advancement in technology and upsurge in network devices, more and more devices are getting connected to the network leading to more data and information on the network which emphasizes the security of the network to be of paramount importance. Malicious traffic must be detected in networks and machine learning or more precisely deep learning (DL), which is an upcoming approach, should be used for better detection. In this paper, Detection of attacks through a classification of traffic into normal and attack data is done using 1D-CNN, a special variant of convolutional neural network (CNN). For this, the CICIDS2017 dataset consisting of 14 attack types spread across 8 different files, is considered for evaluating model performance and various indicators like recall, precision, F1-score have been utilized. Separate 1D-CNN based DL models were built on individual sub-datasets as well as on combined datasets. Also, an evaluation of the model is done by comparing it with an artificial neural network (ANN) model. Experimental results have demonstrated that the proposed model has performed better and shown great capability in detecting network attacks as the majority of the class labels had achieved excellent scores in each of the evaluation indicators used.

Author 1: Kuljeet Singh Author 2: Amit Mahajan

Author 3: Vibhakar Mansotra

Keywords: 1D-CNN; CICIDS2017; network attacks; deep learning

DOWNLOAD PDF

Paper 70: Applying Grey Systems and Inverse Distance Weighted Method to Assess Water Quality from a River

Abstract: The Cañete River basin, in Peru, has suffered an increase in pollution due to various causes, among the main ones being the lack of knowledge, culture of individuals and municipal authorities, economic activities, among others. We analyze this degree of pollution reflected in the upper part of the Cañete river basin through the Grey Clustering method, based on grey systems, which is presented as a good alternative to evaluate water quality in a comprehensive way, making use of Historical data from the monitoring program for the years 2014 and 2015 with nine monitoring points carried out by the National Water Authority (ANA), 6 parameters were defined to evaluate: Hydrogen potential (pH), Biochemical oxygen demand, Chemical demand of Oxygen, Total Suspended Solids, Total Manganese and Total Iron based on the PRATI index. For the spatial distribution, interpolation surfaces of the clustering coefficients were created, using the Spatial Analyst extension of the ArcGIS software, which provides tools to create, analyze and map data in raster format or surfaces. The interpolation method used is Inverse Distance Weighted (IDW). The results of the evaluation showed that in 2014 the monitoring points determined, through the Grey Clustering method, a level of contamination "Uncontaminated" at each point except for point P7 which gives us an "Acceptable" level according to the PRATI indices, while for the year 2015 points P1 and P2 indicate a level of contamination "Moderately contaminated", point 3 an "Acceptable" level, after points P4 to P9 they present a level of "Not contaminated". Finally, the Grey Clustering analysis method will determine the water quality in the 9 monitoring points of the upper-middle

basin of the Cañete River in the years 2014 and 2015. Allowing to observe the reduction of water quality in points P1 and P2 for the period of the years 2014 and 2015 respectively, being crucial to achieve water resource management among local governments that can insert awareness and sustainable development policies.

Author 1: Alexi Delgado Author 2: Anthonny Fernandez

Author 3: Eduardo Lozano Author 4: Dennis Miguel

Author 5: Félix León Author 6: Jhosep Arteta

Author 7: Ch. Carbajal

Keywords: Grey systems; inverse distance weighted (IDW); water quality

DOWNLOAD PDF

Paper 71: Unsupervised Machine Learning Approach for Identifying Biomechanical Influences on Protein-Ligand Binding Affinity

Abstract: Drug discovery is incredibly time-consuming and expensive, averaging over 10 years and $985 million per drug. Calculating the binding affinity between a target protein and a ligand through Virtual Screening is critical for discovering viable drugs. Although supervised machine learning (ML) can predict binding affinity accurately, models experience severe overfitting due to an inability to identify informative properties of protein-ligand complexes. This study used unsupervised ML to reveal underlying protein-ligand characteristics that strongly influence binding affinity. Protein-ligand 3D models were collected from the PDBBind database and vectorized into 2422 features per complex. Principal Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding (t-SNE), K-Means Clustering, and heatmaps were used to identify groups of complexes and the features responsible for the separation. ML benchmarking was used to determine the features' effect on ML performance. The PCA heatmap revealed groups of complexes with binding affinity of pKd<6 and pKd>8 and identified the number of CCCH and CCCCCH fragments in the ligand as the most responsible features. A high correlation of 0.8337, their ability to explain 18% of the binding affinity's variance, and an error increase of 0.09 in Decision Trees when trained without the two features suggests that the fragments exist within a larger ligand substructure that significantly influences binding affinity. This discovery is a baseline for informative ligand representations to be generated so that ML models overfit less and can more reliably identify novel drug candidates. Future work will focus on validating the ligand substructure's presence and discovering more informative intra-ligand relationships.

Author 1: Arjun Singh

Keywords: Drug discovery; unsupervised machine learning; feature engineering; protein-ligand binding affinity; virtual screening

DOWNLOAD PDF

Paper 72: Statistical Analysis of Cybersecurity Awareness Issues in Higher Education Institutes

Abstract: Now-a-days, computers and the Internet are becoming increasingly indispensable tools in several aspects of our lives, including academic study, professional work, entertainment, and communication. Despite the significant advantages of information technology, particularly in information accessibility and internet applications, cyber security has risen to become a national concern in Saudi Arabia, and cyber security threats now need to be taken more seriously. Therefore, computer and network security are a concern not only for traditional security awareness organisations, for example, military, bank, or financial institutions, but also for every individual and government official who use computers. Besides, nowadays, more and more organisations' valuable assets are stored in the computerised information system; security has become an essential and urgent issue. However, it is remarkable that most systems today are designed with little attention to security concerns. This study aims to examine and analyse cyber security issues, including cyber risk, cyber security, cyber security awareness, and cyber trust, among higher education students in Saudi Arabia. Based on an analysis of the collected data using SPSS, the findings of this study highlight a lack of awareness of basic information related to cyber security among Saudi students. In addition, the number of students attending training programs was very low. Considering other security issues, this study reveals that while Saudi students are aware of cyber risk, they are not aware of cyber security. In addition, Saudi students are not aware of and do not have cyber trust.

Author 1: Latifa Alzahrani

Keywords: Cyber security; cyber security awareness; educational institutes; cyber risk; higher education; cyber trust

DOWNLOAD PDF

Paper 73: Ontology-based Daily Menu Recommendation System for Complementary Food According to Nutritional Needs using Naïve Bayes and TOPSIS

Abstract: Babies begin to be given complementary feeding at the age of 6 to 24 months. Complementary foods given to babies need to meet nutritional needs according to their ages. Since, at these ages, babies are just learning to eat, it is necessary to plan a complementary food menu referring to the nutritional needs and the baby and mother's preferences. It is certainly not an easy thing for a mother. Therefore, a recommendation system is needed to determine the baby's daily menu according to those all. This research proposes a complementary food menu recommendation system that considers the balanced composition of three significant nutrients (carbohydrates, protein, and fat) in the diet. It also takes into account the baby and mother's preferences. The ontology contains Knowledge-based about food and its nutritional content and the nutritional needs of babies according to their ages. Naive Bayes is used to prepare menu options according to user preferences. TOPSIS method is used in this study to provide optimal recommendations regarding nutritional balance and user preferences. Several mothers who have had babies aged 6-24 months and mothers of babies aged 6-24 months were asked to test the recommendation system. The results of the usability testing of the system using SUS showed a good level of user satisfaction.

Author 1: Mujahidah Showafah Author 2: Sari Widya Sihwi

Author 3: Winarno

Keywords: Calorie; complementary food; babies; Naïve Bayes; nutrition needs; ontology; recommendation system; SUS; TOPSIS

DOWNLOAD PDF

Paper 74: Emerging Requirement Engineering Models: Identifying Challenges is Important and Providing Solutions is Even Better

Abstract: Requirement Engineering is one of the most crucial tasks because it serves as the foundation for any software. Four pillars of requirement engineering procedures underpin the entire software. The bricks that make up the software edifice are functional and non-functional needs. Finally, design, implementation, and testing add stories to the foundation, allowing a full software tower to be built on top of it. As a result, the foundation must be strong enough to support the remainder of the software tower. Requirement engineers have various hurdles to design successful software for this purpose. Requirement Engineering (RE) is emerging as an increasingly important discipline to promote the development of web applications, as these are designed to meet various stakeholder requirements, additional functional, information, multimedia, and usability requirements compared to traditional software applications. The requirements of software systems are a very important area in software engineering. The success of software systems depends on how it effectively meets the requirements of users. In this paper, the review of current state of requirements engineering in which requirements from users are checked analyzed with their consistency and correctness is presented, and then identifies the emerging models of requirement engineering. Firstly, the paper highlight the current activities that enable the understanding of goals and objectives for developing proposed software systems, then the focus is on the techniques for improving the precision, accuracy, and variety of requirements. Next, identification of the challenges of emerging requirement engineering models is explained. The challenges like security and global trend that posed by emerging models of the future. Finally, we are trying to suggest some solutions for the mentioned challenges.

Author 1: Hina Noor Author 2: Maheen Tariq

Author 3: Anam Yousaf Author 4: Hafiz Wajid Ali

Author 5: Arqam Abdul Moqeet Author 6: Abu Bakar Hamid

Author 7: Mahek Hanif Author 8: Huma Naz

Author 9: Nabeel Tariq Author 10: Prof. Dr. Ijaz Amin

Author 11: Osama Naseer

Keywords: Requirement engineering; current requirement engineering methods; emerging models for requirement engineering; challenges; pros and cons

DOWNLOAD PDF

Paper 75: An Advanced Ontology based Deep Learning for Computer-aided Interpretation of Mammography Images

Abstract: Computer assisted detection (CAD) are able to detect and characterize suspicious mammographic images, micro calcifications, masses or more difficult, architectural distortion. With the exploitation of these different characteristics, the system can specify and predict the severity of the tumor to assess the risk in terms of Malignity/Benignness. Our work involves the development of a new method for screening breast cancer, this is achieved by developing a whole strategy of knowledge extraction through deep learning and medical ontology appropriate for the classification of regions selected from digital mammograms, for each radiological sign considered, namely, masses and micro calcifications. First, we extracted the parameters characterizing the images used as input to a deep convolutional neuron network CNN. The learning is supervised because the images used are images from the MARATHON database of the University of Florida; they are already diagnosed by experts. The second phase aims to add a semantic level to our classification through a specialized ontology developed for this purpose based on the BIRADS characterization system. Based on the evaluation performed, the proposed approach provides better classification results than the usual methods for assisting in the computer aided detection of breast cancer.

Author 1: Hamida Samiha Rahli Author 2: Nacéra Benamrane

Keywords: Computer assisted detection (CAD); breast cancer; the digital database for screening mammography (DDSM); ontology; deep learning; breast imaging-reporting and data system (BIRADS)

DOWNLOAD PDF

Paper 76: Electronic Commerce Product Recommendation using Enhanced Conjoint Analysis

Abstract: While finding any product, there are many identical products sold in the marketplace, so buyers usually compare the items according to the desired preferences, for example, price, seller reputation, product reviews, and shipping cost. From each preference, buyers count subjectively to make a final decision on which product is should be bought. With hundreds of thousands of products to be compared, the buyer may not get the product that meets his preferences. To that end, we proposed the Enhanced Conjoint Analysis method. Conjoint Analysis is a common method to draw marketing strategy from a product or analyze important factors of a product. From its feature, this method also can be used to analyze important factors from a product in the marketplace based on price. We convert importance factor percentage as a coefficient to calculate weight from every attributes and summarize it. To evaluate this method, we compared the ECA method to another prediction algorithm: generalized linear model (GLM), decision tree (DT), random forest (RF), gradient boosted trees (GBT), and support vector machine (SVM). Our experimental results, ECA running time is 6.146s, GLM (5.537s), DT (1s), RF (10,119s), GBT (45.881s), and SVM (11.583s). With this result, our proposed method can be used to create recommendations besides the neural network or machine learning approach.

Author 1: Andrew Brian Osmond Author 2: Fadhil Hidayat

Author 3: Suhono Harso Supangkat

Keywords: Enhanced conjoint analysis; marketplace; ecommerce

DOWNLOAD PDF

Paper 77: Normalisation of Indonesian-English Code-Mixed Text and its Effect on Emotion Classification

Abstract: Usage of code-mixed text has increased in re-cent years among Indonesian internet users, who often mix Indonesian-language with English-language text. Normalisation of this code-mixed text into Indonesian needs to be performed to capture the meaning of English parts of the text and process them effectively. We improve a state-of-the-art code-mixed Indonesian-English normalisation system by modifying its pipeline modules. We further analyse the effect of code-mixed normalisation on emotion classification tasks. Our approach significantly improved on a state-of-the-art Indonesian-English code-mixed text normal-isation system in both the individual pipeline modules and the overall system. The new feature set in the language identification module showed an improvement of 4.26% in terms of F1 score. The combination of machine translation and ruleset in the lexical normalisation module improved BLEU score by 25.22% and lowered WER by 62.49%. The use of context in the translation module improved BLEU score by 2.5% and lowered WER by 8.84%. The effectiveness of the overall pipeline normalisation system increased by 32.11% and 33.82%, in terms of BLEU score and WER, respectively. Code-mixed normalisation also improved the accuracy of emotion classification by up to 37.74% in terms of F1 score.

Author 1: Evi Yulianti Author 2: Ajmal Kurnia

Author 3: Mirna Adriani Author 4: Yoppy Setyo Duto

Keywords: Code-mixed normalisation; Indonesian; English; emotion classification

DOWNLOAD PDF

Paper 78: Sign Language Gloss Translation using Deep Learning Models

Abstract: Converting sign language to a form of natural language is one of the recent areas of the machine learning domain. Many research efforts have focused on categorizing sign language into gesture or facial recognition. However, these efforts ignore the linguistic structure and the context of natural sentences. Traditional translation methods have low translation quality, poor scalability of their underlying models, and are time-consuming. The contribution of this paper is twofold. First, it proposes a deep learning approach for bidirectional translation using GRU and LSTM. In each of the proposed models, Bahdanau and Luong's attention mechanisms are used. Second, the paper experiments proposed models on two sign languages corpora: namely, ASLG-PC12 and Phoenix-2014T. The experiment con-ducted on 16 models reveals that the proposed model outperforms the other previous work on the same corpus. The results on the ASLG-12 corpus, when translating from text to gloss, reveal that the GRU model with Bahdanau attention gives the best result with ROUGE (Recall-Oriented Understudy for Gisting Evaluation) score 94.37% and BLEU (Bilingual Evaluation Understudy)-4 score 83.98%. When translating from gloss to text, the results also show that the GRU model with Bahdanau attention achieves the best result with ROUGE score 87.31% and BLEU-4 66.59%. On Phoenix-2014T corpus, the results of text to gloss translation show that the GRU model with Bahdanau attention gives the best result in ROUGE with a score of 42.96%, while the GRU model with Luong

attention gives the best result in BLEU-4 with 10.53%. When translating from gloss to text, the results report that the GRU model with Luong attention achieves the best result in ROUGE with a score of 45.69% and BLEU-4 with a score of 19.56%.

Author 1: Mohamed Amin Author 2: Hesahm Hefny

Author 3: Ammar Mohammed

Keywords: Sequence to sequence model; neural machine trans-lation; sign language; deep learning; LSTM; GRU

DOWNLOAD PDF

Paper 79: Swapping-based Data Sanitization Method for Hiding Sensitive Frequent Itemset in Transaction Database

Abstract: Sharing a transaction database with other parties for exploring valuable information becomes more recognized by business institutions, i.e., retails and supermarkets. It offers various benefits for the institutions, such as finding customer shopping behavior and frequently bought items, known as frequent itemsets. Due to the importance of such information, some institutions may consider certain frequent itemsets as sensitive information that should be kept private. Therefore, prior to handling a database, the institutions should consider privacy preserving data mining (PPDM) techniques for preventing sen-sitive information breaches. Presently, several PPDM methods, such as item suppression-based methods and item insertion-based methods have been developed. Unfortunately, the methods result in significant changes to the database and induce some side effects such as hiding failure, significant data dissimilarity, misses cost, and artificial frequent itemset occurrence. In this paper, we propose a swapping-based data sanitization method that can hide the sensitive frequent itemset while at the same time it can minimize the side effects of the data sanitization process. Experimental results indicate that the proposed method outperforms existing methods in terms of minimizing the side effects.

Author 1: Dedi Gunawan Author 2: Yusuf Sulistyo Nugroho

Author 3: Maryam

Keywords: Transaction database; data sanitization; data min-ing; sensitive frequent itemset; swapping-based method

DOWNLOAD PDF

Paper 80: Software Security Static Analysis False Alerts Handling Approaches

Abstract: False Positive Alerts (FPA), generated by Static Analyzers Tools (SAT), reduce the effectiveness of the automatic code review, letting them be underused in practice. Researchers conduct a lot of tests to improve SAT accuracy while keeping FPA at a lower rate. They use different simulated and production datasets to validate their proposed methods. This paper surveys recent approaches dealing with FPA filtering; it compares them and discusses their usefulness. It also studies the used datasets to validate the identified methods and show their effectiveness to cover most program defects. This

study focuses mainly on the security bugs covered by the datasets and handled by the existing methods.

Author 1: Aymen Akremi

Keywords: Software security; static analysis; false alert reduc-tion; source code dataset; security bugs

DOWNLOAD PDF

Paper 81: CDRA: A Community Detection based Routing Algorithm for Link Failure Recovery in Software Defined Networks

Abstract: The increase in size and complexity of the Internet has led to the introduction of Software Defined Networking (SDN). SDN is a new networking paradigm that breaks the limitations of traditional IP networks and upgrades the current network infrastructures. However, like traditional IP networks, network failures may also occur in SDN. Multiple research studies have discussed this problem by using a variety of techniques. Among them is the use of the community detection method is one of the failure recovery technique for SDN. However, this technique have not considered the specific problem of multiple link multi-community failure and inter-community link failure scenarios. This paper presents a community detection-based routing algorithm (CDRA) for link failure recovery in SDN. The proposed CDRA scheme is efficient to deal with single link intra-community failure scenarios and multiple link multi-community failure scenarios and is also able to handle the inter-community link failure scenarios in SDN. Extensive simulations are performed to evaluate the performance of the proposed CDRA scheme. The simulation results depicts that the proposed CDRA scheme have better simulations results and reduce average round trip time by 35.73%, avg data packet loss by 1.26% and average end to end delay 49.3% than the Dijkstra based general recovery algorithm and also can be used on a large scale network platform.

Author 1: Muhammad Yunis Daha

Author 2: Mohd Soperi Mohd Zahid Author 3: Babangida Isyaku

Author 4: Abdussalam Ahmed Alashhab

Keywords: Software Defined Network (SDN); community detection methods; CDRA; link failure

DOWNLOAD PDF

Paper 82: A Reliable Lightweight Trust Evaluation Scheme for IoT Security

Abstract: The rapid development of smart devices and the consequent demand their reliability have posed many challenges limiting their versatility. One of the most significant challenges is safeguarding the widespread network of sensors and devices within harsh remote environments. Numerous trust schemes have been proposed to overcome related IoT security concerns. How-ever, most of these schemes are not lightweight and consequently are not energy-efficient. This paper proposes a reliable lightweight trust evaluation scheme (RTE) to mitigate the malicious behavior of the nodes within IoT networks. The nodes

are grouped into a set of clusters each having a cluster head while cluster members are categorized by evaluating their associated residual energy. Nodes with residual energy lower than the threshold (which is determined by the base station) are suspended until they recover and regain their activity. The computations are handled by the CH which is elected by an algorithm according to its energy and coverage degree in order to optimize the energy consumption in the network. For validation and performance evaluation, the proposed RTE scheme was compared to three of the recent schemes in its category. The obtained results have revealed that the proposed RTE scheme outperforms all of them in terms of detection rate, trust evaluation time, and energy efficiency.

Author 1: Hamad Aldawsari Author 2: Abdel Monim Artoli

Keywords: IoT security; clustering; trust; energy efficient algorithm

DOWNLOAD PDF

Paper 83: Multi-Criteria Decision-Making Approach for Selection of Requirements Elicitation Techniques based on the Best-Worst Method

Abstract: During requirements elicitation stage, requirements engineers gather system requirements and drive stakeholders to convey needs and desired software functionality. The elicitation techniques used to acquire software requirements significantly impact the quality of elicited requirements. Several elicitation techniques have been proposed for the Requirement Engineering (RE) process; however, these techniques are rarely used in reality due to a lack of empirical and relative appraisals to assist software team members in deciding on the most appropriate technique. Re-quirement engineers encounter difficulty in deciding the suitable elicitation technique to adopt for a certain software project. This difficulty is due to a lack of knowledge regarding the available elicitation techniques, their efficacy, and how appropriate they are for a certain project. According to the literature, requirements engineering processes benefit from the use of Multi-Criteria Decision-Making (MCDM) approaches within particular contexts. An optimal structure is constitutionally presented within the area of the requirements engineering process; hence, the demonstra-tion of a robust decision-making method in the requirements engineering process should motivate a higher level of satisfaction with software projects developed in this way. This study proposes an approach for using the MCDM method in the requirements engineering process. The study contains a model for investigating the selection of an appropriate elicitation technique based on a decision-making method, namely, the Best-Worst Method (BWM). The findings of the proposed model demonstrate the BWM's power in solving complex decision problems involving several criteria and alternatives.

Author 1: Abdulmajeed Aljuhani

Keywords: Requirements elicitation; elicitation techniques; de-cision support methods; Best-Worst Method; BWM

DOWNLOAD PDF

Paper 84: e-Business Model to Optimise Sales through Digital Marketing in a Peruvian Company

Abstract: The COVID-19 pandemic has affected the Peruvian market, generating a great loss in sales in Peruvian companies. The objective of the research is to develop a model to optimize sales with the use of digital marketing in a Peruvian company, the chosen methodology is DesingScrum, which is a hybrid of Scrum and Desing Thinking, with 10 phases (empathize, define. It has 10 phases (empathise, define, ideate, planning meeting, sprint backlog, daily meeting, sprint review, sprint retrospective sprint, prototype and testing) and the MarvelApp tool was used to create the prototype. The results are obtained after the completion of the review of each sprint, showing in detail how it was progressing in each sprint, and through the retrospective evaluated the development of the project for the realization of continuous improvement in the next product. Further prototype was made, with the application MarvelApp, which shows the model of e-business. Then testing was done through a survey that customers gave their opinions about the prototype and finally the digital marketing proposal was made by a model, which explains the interconnected tools to attract new customers. The conclusion is the construction of the digital marketing model according to the needs of the context to improve the sales of the company through e-business.

Author 1: Misael Lazo-Amado Author 2: Leoncio Cueva-Ruiz

Author 3: Laberiano Andrade-Arenas

Keywords: Design thinking; desingScrum methodology; digital marketing; e-Business; scrum

DOWNLOAD PDF

Paper 85: A Fast and Efficient Algorithm for Outlier Detection Over Data Streams

Abstract: Outlier detection over data streams is an important task in data mining. It has various applications such as fraud detection, public health, and computer network security. Many approaches have been proposed for outlier detection over data streams such as distance-,clustering-, density-, and learning-based approaches. In this paper, we are interested in the density-based outlier detection over data streams. Specifically, we propose an improvement of DILOF, a recent density-based algorithm. We observed that the main disadvantage of DILOF is that its summarization method has many drawbacks such as it takes a lot of time and the algorithm accuracy is significant degradation. Our new algorithm is called DILOFC that utilizing an efficient summarization method. Our performance study shows that DILOF outperforms DILOF in terms of total response time and outlier detection accuracy.

Author 1: Mosab Hassaan Author 2: Hend Maher

Author 3: Karam Gouda

Keywords: Data mining; outlier detection; data streams; density-based approach; clustering-based approach

DOWNLOAD PDF

Paper 86: Network Intrusion Detection System based on Generative Adversarial Network for Attack Detection

Abstract: The Intrusion Detection System (IDS) is the main element to prevent malicious traffic on the network. IDS will quickly increase the ability to detect network threats with the help of Deep Learning algorithms. As a result, attackers are finding new ways to evade identification. Polymorphic attacks, search for the attackers, as they can bypass the IDS. Generative Adversarial Networks (GAN) is a method proven in generating various forms of data. It is becoming popular among security researchers as it can produce indistinguishable data from the original data. This work proposed a model to generate DDoS attacks using a GAN. Several techniques have been used to regenerate the feature selection to identify the attack and generate polymorphic data. The data will change feature profile in every cycle to test if the IDS can detect the new version of attack data. Simulation results from the proposed model show that with constant changing attack profiles, defending arrangements that handle incremental knowledge will yet stay exposed to current attacks.

Author 1: Abhijit Das Author 2: S G Balakrishnan

Author 3: Pramod

Keywords: Generative adversarial networks; network threats; deep learning; intrusion detection; feature selections

DOWNLOAD PDF

Paper 87: Near-ground Measurement and Modeling for Archaeological Park of Pisac in Cusco for LoRa Technology

Abstract: This research work presents the details of a near-ground received power measurement and modeling based on it, inside the archaeological park of Pisac in the city of Cusco. The measurements were performed at a working frequency of 920 MHz in the industrial, scientific, and medical (ISM) band, using transceiver devices with LoRa technology. The power of the received signal is obtained while the transmitter moves at a constant speed of 0.4 m/s, to characterize the fading that occurred within this archaeological park, making appropriate use of the moving average filter separates the fading to large and small scale, then on the filtered signal is used the algorithm of linear regression, to obtain a model that characterizes the exponent of propagation loss and the shading factor. In addition, the small-scale fading is characterized according to its proba-bility distribution, the statistical parameters of the distributions are obtained based on the small-scale measurements. We also measured variations of receiving antenna height, ranging from 20cm, 50cm, 80cm, and 120cm above the ground, finding that the height has a strong influence on the propagation loss exponent, while the shadow variation shows a smaller variation. The model obtained is validated by the coefficient of determination and the root mean square error (RMSE) value.

Author 1: Yhon D. Lezama Author 2: Jinmi Lezama

Author 3: Jorge Arizaca-Cusicuna

Keywords: Archaeological park; LoRa; propagation model; statistics; near-ground

DOWNLOAD PDF

Paper 88: Development of Predictions through Machine Learning for Sars-Cov-2 Forecasting in Peru

Abstract: The SARS-COV-2 virus of the coronavirus family was identified in 2019. This is a type of virus that infects humans and some animals, in Peru it has seriously affected everyone, causing so many deaths, which has resulted in that people be tested to rule out contagion, using laboratory methods recommended by the government of the country. Therefore, the data science methodology was used with this research, where its objective is to predict what types of people are contaminated during SARS-COV-2 by the regions of Peru, identified through laboratory methods, therefore, the ”data bank” was taken by PNDA, the CSV file was used for that study, apart from the fact that it comes from the INS and the CDC of the MINSA. In which, machine learning was developed with the decision tree algorithm and then began coding, in such a way that the distribution called Anaconda was used where it is encoded in Python language, together with that distribution, Jupyter Notebook was used which is a client-server application. The results generated by this research prove that it was possible to identify the types of individuals by SARS-COV-2. These results can help prevention entities against SARS-COV-2 to apply the corresponding preventive measures in a more focused way.

Author 1: Shal´om Adonai Huaraz Morales

Author 2: Marissel Fabiola Mio Antayhua

Author 3: Laberiano Andrade-Arenas

Keywords: Forecast; laboratory methods; machine learning; Python; SARS-COV-2

DOWNLOAD PDF

Paper 89: Inclusive Education: Implementation of a Mobile Application for Blind Students

Abstract: Currently, the world is going through an era of changes in the education sector, but most Latin American countries are lagging, especially Peru, which does not have the technological tools that allow it to advance to an adequate level of inclusion of disabled students, especially blind students who are 60% of students who drop out of school for lack of education that have be suitable for their need. Consequently, the present research work is originated which aims to develop a mobile application oriented to the benefit of inclusive education of blind students. Therefore, the agile scrum methodology was used for the development of this project, executed in 5 phases, the requirements identified for the development of the mobile application were obtained through a questionnaire to 25 parents of visually impaired students, allowing the development of a mobile application that meets quality of inclusive education that can be applied in the education sector. Finally, as a result of the research work, another satisfaction survey was conducted with 50 parents where the application was evaluated, obtaining 90%of acceptance and satisfaction.

Author 1: Alejandro Boza-Chua

Author 2: Karen Gabriel-Gonzales

Author 3: Laberiano Andrade-Arenas

Keywords: Blind; disability; educational inclusion; mobile ap-plication; scrum methodology

DOWNLOAD PDF

Paper 90: Design and Implementation of HSQL: A SQL-like language for Data Analysis in Distributed Systems

Abstract: In today's modern world, we're experiencing a substantial increase in the use of data in various fields, and this has necessitated the use of distributed systems to consume and process Big Data. Machine learning tends to benefit from the usage of Big Data, and the models generated from such techniques tend to be more effective. However, there is a steep learning curve to getting used to handling Big Data, as traditional data management tools fail to perform well. Distributed systems have become popular, where the task of data processing is split amongst various nodes in clusters. SQL, is a popular database management language popular to data scientists. It is often given second class support, where SQL can be embedded into a primary language of use (e.g. SQL in Scala for Spark), which allows for using SQL but one still needs to know the primary language of the platform (Scala, as per the example, or ECL in HPCC Systems). It may also be present as a supported language. In either case, using useful tooling such as Visualizing data and creating and using machine learning models become difficult, as the user needs to fall back to the primary language of the system. In the proposed work, a new SQL-like language, HSQL, an open source distributed systems solution, was developed for allowing new users to get used to its distributed architecture and the ECL language, with which it primarily operates with (which was chosen as a target). Additionally, a program that could translate HSQL-based programs to ECL for use was made. HSQL was made to be completely inter-compatible with ECL programs, and it was able to provide a compact and easy to comprehend SQL-like syntax for performing general data analysis, creation of Machine learning models and visualizations while allowing a modular structure to such programs.

Author 1: Anurag Singh Bhadauria Author 2: Atreya Bain

Author 3: Jyoti Shetty Author 4: Shobha G

Author 5: Arjuna Chala Author 6: Jeremy Clements

Keywords: ANTLR4; big data; context free grammar; dis-tributed systems; HPCC; Javascript; language; machine learning; Parser; SQL; transpiler

DOWNLOAD PDF

Paper 91: Factors Influencing the Adoption of Cyber Security Standards Among Public Listed Companies in Malaysia

Abstract: Employee's failure to adhere to their organization's cyber security policies contributes most of the cyber incidents. To secure information security systems, companies need to communicate behavioral and technical solutions to their employees, due to the fragility of the human factor since it plays a critically significant role in securing cyber systems. The necessity to safeguard information systems have speed up the evolution of the present method of cyber security, which should be based on adequately adopting cyber security standards to secure business enterprise's assets and users in cyberspace. This paper studies factors influencing the adoption of cyber security standards

among public listed companies in Malaysia. Through online survey that was distributed among 275 Public listed companies. The findings indicated that expected related benefits and perceived ease of use had significant impact on the adoption of cyber security standards. On the other hand, perceived security had played important moderating influence on the relationship between organizational factors and the adoption of cyber security standards.

BACK TO TOP

COMPUTER SCIENCE JOURNAL

About the Journal

Call for Papers

Submit Paper

Indexing

OUR CONFERENCES

Computing Conference

Intelligent Systems Conference

Future Technologies Conference

Communication Conference

HELP & SUPPORT

Contact Us

About Us

Terms and Conditions

Privacy Policy

Machine Learning for Diagnosing Drug Users and Types of Drugs Used

Anthony Anggrawan[1], Ni Gusti Ayu Dasriani[5], Mayadi[6]
Computer Science Study Program
Bumigora University, Mataram, Indonesia

Che Ku Nuraini[3]
Department of Centre of Research and Innovation Management, Universiti Teknikal Malaysia Melaka, Malaysia

Christofer Satria[2]
Visual Communication Design Study Program
Bumigora University, Mataram, Indonesia

Lusiana[4]
Technical Information Study Program
STMIK AMIK Riau, Riau, Indonesia

***Abstract*—Drug use is very detrimental to the physical and psychological health of users. Drug abuse also causes addiction and is a global epidemic. Therefore it is not surprising that scientific research related to drugs has attracted attention for research. However, many factors become obstacles in the medical services of the drug user, including cost, flexibility, and a slow process. Meanwhile, electronic systems can speed up handling time, improve work efficiency, save costs and reduce inspection errors. It means that a breakthrough is needed in developing a platform that can identify drug users. Therefore, this research aims to build machine learning with expertise like an expert who can diagnose drug users and distinguish the types of drugs used by drug users. The expert system on machine learning was developed using the Forward Chaining and Certainty Factor methods. This study concludes that the expert system on machine learning developed can be used to diagnose drug users and distinguish the types of drugs used with an accuracy of up to 80%. The implications of the expert system on machine learning are an alternative method for narcotics officers and medical doctors in diagnosing drug users and the types of drugs used.**

***Keywords*—Machine learning; drug; expert system; forward chaining; certainty factor**

I. Introduction

Social environment factors have influenced others to engage in drug use [1] [2]. Adult figures who are addicted to drugs have a great influence on the behavior of others to become addicted [3]. Poor, low skills, life pressure, anxiety, and deviant behavior are factors that also lead to drug use [3]. Relaxing, drinking, staying up late, increasing enthusiasm, and relieving stress are other triggering factors for many drug use among young people [4]. Drug use, including amphetamines, marijuana, cocaine, heroin, and the like, are a major public health problem in physiological symptoms, resulting in behavioral changes, cognitive problems, and mental health [5]. Drug use also affects the physiology and behavior of future generations [2].

Drug abuse causes physical dependence (addiction) or relapse to continuously consume [6], although it has resulted in physical and psychological problems [6]. The previous research shows that drug users are very high [7] [1] and increasing globally [8]. Drug abuse has become a global epidemic that affects human behavior [9]. Because drug use is very detrimental to the physical and psychological harm of the user, it is not surprising that this research related to drugs has attracted attention for scientific research [10].

Another factor that is often considered in medical services is the cost factor and its inflexibility (Bevan & Patel, 2016). Processes that are done manually tend to cause delays in medical diagnosis [11]. Using an electronic system can speed up handling time, improve work efficiency, and save costs [11]. Using an electronic system allows lower errors and eliminates omissions in deciding the test results and achieving the results [12]. However, the success of curing drug abuse and dependence is still limited; this includes the lack of success in the early identification of at-risk populations, resulting in increased death rates due to overdose [13]. In other words, not paying attention to the early symptoms of consuming drugs will be disastrous and make people who are loved suffer the destructive effects of the substance [14].

Meanwhile, if it turns out to be able to identify it early, it can prevent harmful consequences in the future that are sure to occur [14]. It means that there is a need for breakthroughs in developing platforms that can identify and screen patients susceptible to addiction after using opioid drugs [13]. Therefore, this research aims to develop a machine learning that has expertise like an expert. The expert system created can identify and screen or diagnose early drug users and the types of drugs used by using the Certainty Factor and Forward Chaining methods. The certainty factor method measures the certainty of the type of drug used by the user or patient who conducts consultations. On the other hand, forward chaining plays a role in the flow of the reasoning process from beginning to end based on data mining of physical symptoms of drug users and the types of drugs used (collected or explored previously).

Medical data is helpful as the knowledge that helps make scientific decisions regarding drug use [15]. Electronic medical use based on doctor's notes is useful for an effective treatment medium [10]. In the meantime, data mining is capable of electronic checks based on the patient's medical record [10].

Besides, the machine learning methods are a technique that can be useful for finding correlations based on the case for prediction purposes [16]. Unfortunately, machine learning is still few in the medical field due to technical problems [17]. Therefore, the simple machine learning method built in this research by imitating (studying) human knowledge in analyzing the physical symptoms of drug users and then implementing it in predicting drug users and identifying drug types used by users. It means that the expert system in machine learning has intelligence like an expert in diagnosing users and the types of drugs used from the physical symptoms that arise from drug users. Taking into account that the current use of information and communication technology (ICT) is growing or expanding very quickly or booming [18]. Therefore, the embodiment of the machine learning system in this research is website-based. So, anyone (the public) can access it from anywhere and has flexibility because it can work on various devices and operating systems. Therefore, a machine learning system in this research is helpful for early diagnosis without having to examine a narcotics laboratory and without a doctor or expert.

It is necessary to know the percentage of machine learning efficacy in identifying drug users and the type of drug used. It means that further testing to determine the actual percentage of machine learning efficacy still needs to be done. This study makes this happen by comparing the test results achieved by machine learning based on symptoms of drug users compared to the test results achieved from laboratory tests of drug users' urine in identifying drug users and the type of drug used. If machine learning has high efficacy, it can save time and cost of drug testing for suspects or drug users by using machine learning compared to testing drugs on urine or blood for suspects and drug users.

Some recent works related to this research:

- Zhongheng Zhang (2016) introduced the k-nearest neighbor (kNN) method as a simple machine learning method for modeling [17]. The similarity between the research in this article and the previous one is that they both use a simple approach to machine learning. While the difference is that the research uses the certainty factor method and forward chaining for machine learning, while previous research uses the kNN method for machine learning. Another difference is that the previous research was focus on predicting the class from the new dataset to the most similar class. In contrast, the research in this article focuses on machine learning to diagnose drug users and the types of drugs used.
- Anthony Anggrawan, Khasnur Hidjah, and Jihadil Qudsi S. (2017) implement intelligent application programs to detect kidney failure [19]. The previous research and the research in this article have similarities in developing web application programs with the PHP programming language and MySQL database. In addition, the last analysis used medical data on failure cases to diagnose new renal illness issues using CBR (Case-Based Reasoning method). In contrast, the articles in this study use the expertise of drug experts (specialists) as knowledge of the application system for early diagnosis of drug users and the types of drugs used by drug users using the Forward Chaining and Certainty Factor methods.
- Kurnia Muludi, Radix Suharjo, Admi Syarif, and Fitria Ramadhani (2018) identified tomato plant diseases [20]. This previous research and the research in this article both implements forward chaining and certainty factor methods. However, the difference in the last research is to build an expert system to identify plant diseases based on android [20]. In contrast, the research in the article builds an expert system to identify users of drugs and the types of drugs used based on the website.
- Munaiseche, Kaparang, and Rompas (2018) built an expert system to assist doctors in diagnosing eye diseases [21]. In contrast to the research in this article, it is to create an expert system to diagnose drug users and the types of drugs used. Furthermore, this previous research used the forward chaining method, while the research in this article uses the forward chaining method and certainty factor. The similarities between the previous study and the research in this article are that both use PHP and MySQL in building an expert system.
- Ninive Von Greiff and Lisa Skogens (2021) investigated a drug user recovery program for drug addiction [22]. The research method is the interview or qualitative approach [22]. The similarity of the research in this article with previous studies is that they both study drug users. The difference is in previous studies examining the results of addiction recovery on drugs with the interview method. Meanwhile, the research in this article builds machine learning that has an intelligent system to detect drug users and the types of drugs used.

The latest related work identifies that the article in this study has a novelty that no previous researcher has researched. Another strength of this research is conducting a comparative test to determine the efficacy of machine learning or expert systems developed in identifying users and the types of drugs used by users that have not been studied before.

The systematics of writing this paper is as follows: the following sub-section discusses the research methodology, which includes research data and research methods used. The next subsection discusses the results and discussion of the research. Finally, the conclusions obtained from the study results and suggestions for further research are narrated in the Conclusions subsection.

II. Research Methodology

This study is a case study at the Indonesian National Narcotics Agency (Badan Narkotika Nasional or BNN) in Mataram, Indonesia. The number of drug users used as samples to test the expertise and accuracy of the machine learning built in this study was 30. The selection of data samples in this study was random. This research's

development of machine learning expertise consists of stages: knowledge acquisition, expert system design (programming), machine learning/expert system testing, and accuracy test (see Fig. 1).

A. Knowledge Acquisition

For the system development stage, the effort made is to obtain knowledge from drug experts, which is used as a knowledge base to build the machine learning or expert systems. The method used in obtaining knowledge related to narcotics is the interview method. The knowledge gained is the knowledge about the types of drugs and their symptoms. Based on the knowledge obtained, there are ten types of drugs that drug users dominantly use, and there are 27 types of drug symptoms.

In the knowledge acquisition stage, an expert from the Indonesian National Narcotics Agency (in Mataram, Indonesia) shares knowledge about drug use, including those related to symptoms and the types of drugs used by drug users. The knowledge gained at this stage serves as a knowledge base in building expertise from machine learning.

B. Expert System Design

The expert system design stage is a process for modeling the data that has been collected and designing an application system that is planned according to programming problems and the acquisition of knowledge obtained. This stage in computer science is known as planning the use case diagram design, data flow diagram (DFD) design, database design, and flowchart to be built on the application program. The programming stage is the implementation stage of the system design plan into a computer programming language [23]. This research uses the PHP and MySQL programming languages. The computer application program that is built is a cloud application program. Application development with PHP programming language and MYSQL database makes application programs can run via the web. By being stored on the server computer, the application program is ubiquitous. The ubiquitous application program means that the application program can be accessed from anywhere and at any time [24].

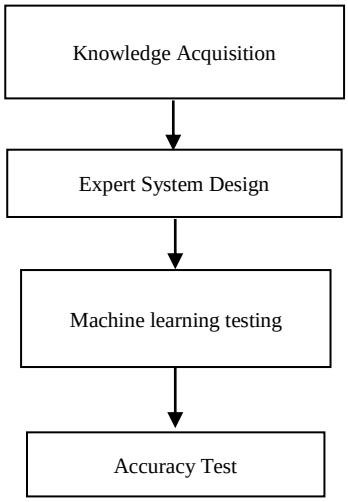

Fig. 1. Stages in the Development of an Expert System on Machine Learning.

C. Machine Learning Testing

The machine learning testing phase is the functional testing phase of the built application or black-box testing. Black-box testing is a test that no longer involves programming code or programming languages. In short, black-box testing on the expert system in this study is to determine whether the expert system built is under the list of desired system requirements.

D. Accuracy Test

Testing accuracy in machine learning is to determine the level of expertise of the expert system built in this study. It means that it is known how much accuracy the expert system has built-in diagnosing users' types of narcotics.

III. Result and Discussion

A. Knowledge Acquisition

The knowledge acquisition stage is the stage of acquiring the required knowledge data. The acquired knowledge acquisition data is useful in solving programming logic in diagnosing users and the types of drugs used by drug users. Table I shows the code for the type of drug used by drug users. Meanwhile, Table II presents the code of symptoms caused by drug users.

TABLE I. List of Types of Drugs

No.	Drug Type Code	Drug Name
1	P01	Cocaine
2	P02	Marijuana
3	P03	Ecstasy
4	P04	Heroin
5	P05	Methamphetamine
6	P06	Hallucinogen
7	P07	Amphetamine
8	P08	Pethidine
9	P09	Codeine
10	P10	Morphine

TABLE II. List of Symptom of Drug user

Symptoms of Drug User			
Code	**Symptom**	**Code**	**Symptom**
G01	Out of breath	G15	Difficult to focus
G02	Anxious and restless	G16	Difficult to rest
G03	Nausea and vomiting	G17	Weight loss
G04	Diarrhea	G18	Dry mouth
G05	Convulsions	G19	Blurred vision
G06	Easy to get angry	G20	Changes in skin color
G07	Depression	G21	Constipation
G08	Changes in sleep patterns	G22	Stomachache
G09	Sweating	G23	Drowsiness
G10	Chills (Hot cold)	G24	Itching
G11	Shaking	G25	Difficulty urinating
G12	Insomnia	G26	Mood swings
G13	Fast heart rate	G27	Dizziness
G14	Increased blood pressure		

TABLE III. RULE BASE OF TYPES AND SYMPTOMS OF BASIC DIAGNOSIS

Rule Base	
Drug Type	**Symptom**
P01	G01 and G02 and G03 and G04 and G05
P02	G06 and G02 and G07 and G08 and G09 and G10
P03	G05 and G11 and G12 and G13 and G14
P04	G15 and G02 and G07 and G16
P05	G11 and G01 and G16 and G17 and G18
P06	G09 and G11 and G18 and G19 and G10 and G14
P07	G18 and G03 and G04 and G05 and G01 and G20
P08	G07 and G13 and G05 and G03
P09	G27 and G03 and G18 and G21 and G22
P10	G023 and G24 and G09 and G25 and G26

After modeling the acquired knowledge acquisition data or knowledge representation (as shown in Table III) the next step is to implement it into the certainty factor algorithm. The certainty factor uses a value between 0.2 and 1.0 to assume a level of confidence in the data. A simulation of the calculation of the certainty factor was carried out based on the weight of symptoms arising from the type of drug used by drug users with weights of 0.8 and 1.0 according to the opinion of the drug expert (see Table IV).

TABLE IV. DETERMINATION OF DRUG SYMPTOM WEIGHT SCORE ACCORDING TO THE DECISION OF DRUG EXPERTS

No	Symptom	Weight Score
1	Very often	1
2	Often	0,8
3	Never	0

So, on the drug symptom weighted score given to the certainty factor, a score of 0 indicates that drug users do not experience these symptoms. If a drug user experiences symptoms, then the weighted score given for the frequently experienced symptoms is 0.8 and the most frequently experienced is 1.0, according to the drug expert's decision.

This study's knowledge base of machine learning expertise is the symptoms, types of drugs, and CF rules obtained from drug experts (see Table V). The knowledge base is an essential component that contains the knowledge possessed by competent experts in the related field (i.e., narcotics in this study). Furthermore, the knowledge base is the basis for decision-making in an expert system, where this decision-making is related to the process of retrieving previously collected and stored knowledge.

B. Expert System Design

Fig. 2 shows the interactions that occur between system users and the developed expert system.

This study has a database that stores records of users, patients, symptoms, and types of drugs, including diagnostic data, so it is necessary to design a data workflow model to realize a structured program.

TABLE V. KNOWLEDGE BASE

Symptom	**J-001**	**J-002**	**J-003**	**J-004**	**J-005**	**J-006**	**J-007**	**J-008**	**J-009**	**J-010**
	CF	**CF**	**CF**	**CF**	**CF**	**CF**	**CF**	**CF**	**CF**	**CF**
Out of breath					0.8		1			
Anxious and restless	0.8	0.8		1						
Nausea and vomiting	0.8						0.8	0.8	0.8	
Diarrhea	1						0.8			
Convulsions	0.8		0.8				0.8	0.8		
Easy to get angry		1								
Depression		0.8		0.8				1		
Changes in sleep patterns		0.8		0.8						
Sweating		0.8				0.8				0.8
Chills		0.8				0.8				
Shaking			0.8		0.8	0.8				
Insomnia			0.8							
Fast heart rate			0.8					0.8		
Increased blood pressure			0.8			1				
Difficult to focus				1						
Difficult to rest					0.8					
Weight loss					0.8					
Dry mouth					1	0.8	0.8		1	
Blurred vision						0.8				
Changes in skin color							0.8			
Constipation									0.8	
Stomachache									0.8	
Drowsiness										1
Itching										0.8
Difficulty urinating										0.8
Mood swings										1
Dizziness									0.8	

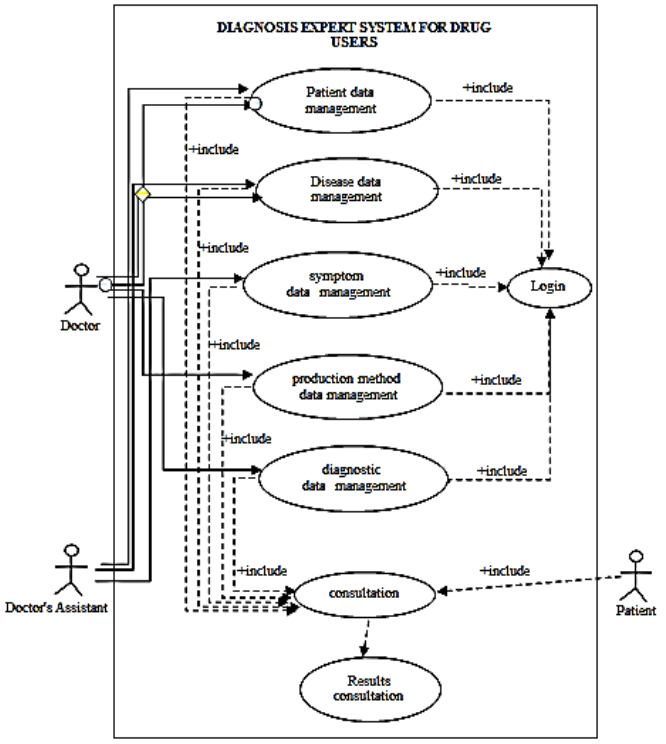

Fig. 2. Use Case Diagram on Machine Learning.

The Data Flow Diagram (DFD) in Fig. 3 and Fig. 4 illustrates where the data flow comes from and where the data processing on the expert system is built.

The context diagram in Fig. 3 shows the data flow of the system globally. In contrast, the overview diagram in Fig. 4 shows a more detailed data flow that the system performs and its engagement with external data.

The flow diagram in Fig. 5 shows a series of flow relationships in the expert system built in this study or shows the overall process sequence in building an expert system in this study.

The flow diagram contains a more detailed description of how each step of the procedure is actually carried out in building an expert system on machine learning that can diagnose users and the types of drugs used by users.

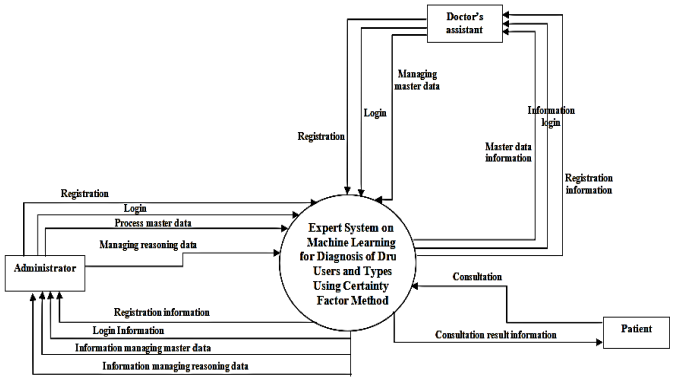

Fig. 3. Context Diagram of Data Flow on Machine Learning.

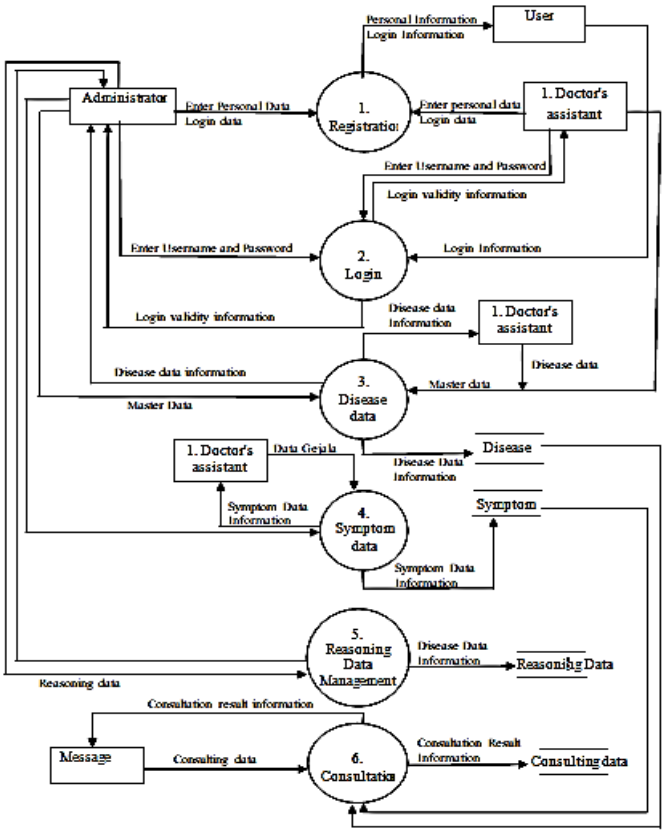

Fig. 4. Overview Diagram of Data Flow on Machine Learning.

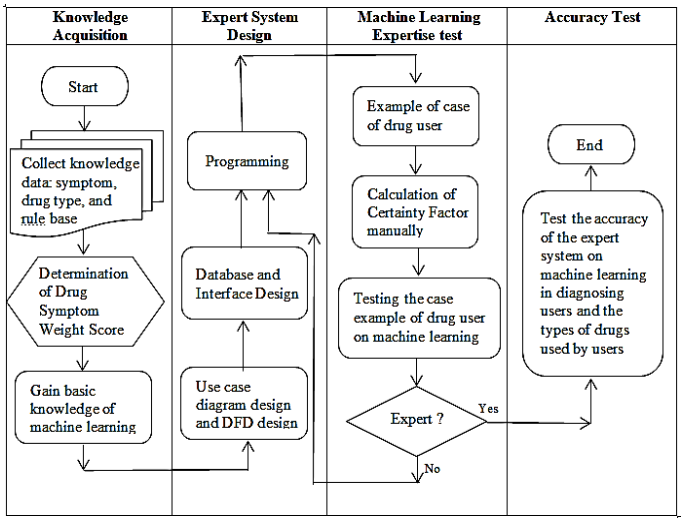

Fig. 5. Flow Diagram of the whole Process of Building an Expert System on the Machine Learning.

C. Machine Learning Testing

Expertise testing of machine learning is carried out using case samples from a patient. For example, in one case, a drug patient had symptoms of shortness of breath, depression, chills, anxiety and restlessness, and irritability. Symptoms of drug patients who have symptoms of shortness of breath, depression, chills, anxiety and restlessness, and irritability are symptoms of drug users: (1) Cocaine, (2) Cannabis, (3) Heroin, and (4) Amphetamine.

The formula for CF is:

CF[H,E] = CF[H] * CF[E]

CF Combine CF[H,E]1 = CF[H,E]1 + CF[H,E] 2 * (1 - CF [H,E] 1)

CF Combine CF[H,E] old3 = CF[H,E] old + CF[H, E] * (1 - CF[H,E] old)

Based on manual calculations, the results are as follows:

1. For J-001 = Cocaine
   G01 = Out of breath (0.8)
   CF[H,E] = CF[H] * CF[E]
   = (0.8 * 0.8)
   = 0.64
   G02 = Anxious and restless (0.8)
   CF[H,E] = CF[H] * CF[E]
   = (0.8 * 0.8)
   = 0.64
   CFk1 = CF[H,E] 1 + CF[H,E]2 * (1 - CF[H,E]1)
   = 0.64 + 0.64 * (1 - 0.64)
   = 0.870
   So the expert CF from the symptoms entered by the user for the type of drug Cocaine is probably 0.870 or 87%.
2. For J-002 = Marijuana
   G07 = Depression (0,8)
   CF[H,E] = CF[H] * CF[E]
   = (0.8 * 0.8)
   = 0.64
   G10 = Chills (0.8)
   CF[H,E] = CF[H] * CF[E]
   = (0.8 * 0.8)
   = 0.64
   G02 = Anxious and restless (0.8)
   CF[H,E] = CF[H] * CF[E]
   = (0.8 * 0.8)
   = 0.64
   G06 = Easy to get angry (1)
   CF[H,E] = CF[H] * CF[E]
   = (0 * 1)
   = 0
   CFk1 = CF[H,E]1 + CF[H,E]2 * (1 - CF[H,E]1)
   = 0.64 + 0.64 * (1 - 0.64)
   = 0.870
   CFk2 = CFk1 + CF[H,E]3 * (1 - CFk1)
   = 0.870 + 0.64 * (1 - 0.870)
   = 0.953
   CFk3 = CFk2 + CF[H,E]4 * (1 – CFk2)
   = 0.953 + 0 * (1 - 0.953)
   = 0.953
   So the CF of the symptoms entered by the user for the type of marijuana drug is likely to be 0.953 or 95%.
3. For J-004 = Heroin
   G07 = Depression (0.8)
   CF[H,E] = CF[H] * CF[E]
   = (0.8 * 0.8)
   = 0.64
   G02 = Anxious and restless (1)
   CF[H,E] = CF[H] * CF[E]
   = (0 * 1)
   = 0
   CFk1 = CF[H,E]1 + CF[H,E]2 * (1 - CF[H,E]1)
   = 0.64 + 0 * (1 - 0.64)
   = 0.64
   So the CF of the symptoms entered by the user for the type of heroin drug is most likely 0.64 or 64%.
4. For J-007 = Amphetamine
   G01 = Out of breath (1)
   CF[H,E]1 = CF[H]1 * CF[E]2
   = (0.8 * 1) = 0.8
   So the CF of the symptoms entered by the user for the type of Amphetamine is most likely 0.8 or 80%.

Based on the value of manual calculations, the largest CF value is taken, which is 0.953 or 95% with the type of marijuana drug. It means the patient is using a type of marijuana drug. A case example is tested on an expert system application program (or on machine learning). If the patient's symptoms in the case sample (with the same symptoms) are entered into the expert system built in this study, the result of the process is as shown in Fig. 6.

Expert system testing on machine learning shows that the expert system has succeeded in correctly identifying the user and the type of drug used by the user. In order to know how accurate the machine learning expertise is, this study also tested several other patients by comparing the results with the urine test results at the Indonesian National Narcotic Agency laboratory in Mataram, Indonesia.

HASIL KONSULTASI

Nama	: Andi
Umur	: 24
Jenis Kelamin	: Laki-laki
Pekerjaan	: PNS
Alamat	: ampenan

No	Pertanyaan	Jawaban
1	Sesak Nafas	TIDAK
2	Cemas dan Gelisah	IYA
3	Mual dan Muntah	TIDAK
4	Diare	TIDAK
5	Kejang-kejang	TIDAK
6	Mudah Marah	TIDAK
7	Depresi	IYA

Fig. 6. Screenshot of Expert System Questions about Drug Symptoms Experienced by Patients.

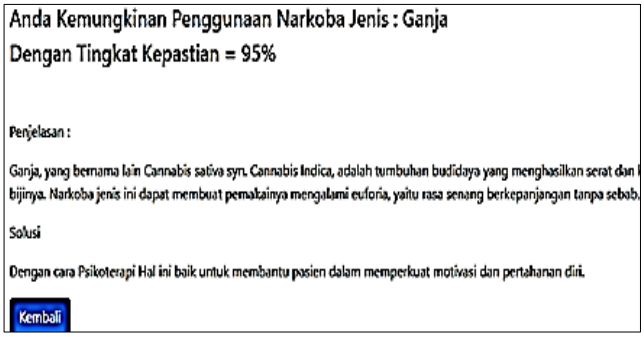

Fig. 7. Screenshot of the Expert System Test Results on the Type of Drug used by the Patient.

Fig. 7 describes it as follows: You are probably using a type of drug with a 95% certainty. Another narrative in Fig. 7 is: another name for cannabis is cannabis sativa. Cannabis sativa is a cultivated plant that contains fiber and narcotic substances in its seeds. This drug makes user experience euphoria, namely a prolonged feeling of pleasure for no reason. The cure is psychotherapy, which helps the patient strengthen the motivation to stop using it.

Expert system testing on machine learning shows that the expert system has succeeded in correctly identifying the user and the type of drug used by the user. In order to know how accurate the machine learning expertise is, this study also tested several other patients by comparing the results with the urine test results at the Indonesian National Narcotic Agency laboratory.

D. Accuracy Test of Machine Learning

The accuracy test of machine learning in this study is to determine the expert performance of the application system built-in diagnosing users and the types of drugs used.

Testing the level of accuracy of machine learning expertise is to compare the suitability of the results with the urine test results from patients at the Indonesian National Narcotics Agency. In a trial of 30 times on 30 patients, there were 30 results of machine learning tests that can correctly identify drug users and as many as 24 machine learning test results that can detect the types of drugs used by drug users. It means that the test results on the data of 30 drug patients show that an expert system on machine learning built using the Certainty Factor method has expertise in diagnosing drug users up to 80 percent. The accuracy rate of up to 80 percent is obtained from the calculation results of 24 divided by 30 and multiplied by 100%.

The details of the machine learning accuracy test results are as described in Table VI. Table VI shows the comparison between the results of the expert system and the results of drug experts on the diagnosis of the types of drugs used by drug users.

TABLE VI. MACHINE LEARNING EXPERTISE ACCURACY TEST RESULT

Number	Case	System Result	Expert Result	Suitability
1	G01, G02, G03, G04, G05	Cocaine	Cocaine	Suitable
2	G06, G02, G07, G08, G09, G10	Marijuana	Marijuana	Suitable
3	G05, G11, G12, G13, G14	Ecstasy	Ecstasy	Suitable
4	G15, G02, G07, G16	Heroin	Heroin	Suitable
5	G11, G01, G16, G17, G18	Methamphetamine	Methamphetamine	Suitable
6	G09, G11, G18, G19, G10, G14	Hallucinogen	Hallucinogen	Suitable
7	G18, G03, G04, G05, G01, G20	Amphetamines	Amphetamines	Suitable
8	G07, G13, G05, G03	Pethidine	Pethidine	Suitable
9	G27, G03, G18, G21, G22	Codeine	Codeine	Suitable
10	G23, G24, G09, G25, G26	Morphine	Morphine	Suitable
11	G03, G06, G07, G15	Pethidine	Pethidine	Suitable
12	G01, G08, G09, G18	Codeine	Codeine	Suitable
13	G01, G02, G05, G09, G11,G15, G18	Codeine	Hallucinogen	Not suitable
14	G04, G12, G13, G17	Ecstasy	Ecstasy	Suitable
15	G08, G11, G17, G18, G19	Methamphetamine	Methamphetamine	Suitable
16	G01, G13, G15, G20, G22	Amphetamines	Amphetamines	Suitable
17	G17, G20, G21, G22, G23, G25, G27	Codeine	Morphine	Not suitable
18	G02, G07, G11, G15, G16	Pethidine	Heroin	Not suitable
19	G03, G07, G10, G14, G19	Pethidine	Pethidine	Suitable
20	G08, G11, G15, G18, G19, G20, G22, G23	Codeine	Amphetamines	Not suitable
21	G02, G06, G07, G08, G09, G10, G19, G27	Marijuana	Marijuana	Suitable
22	G05, G11, G12, G14, G19, G20	Hallucinogen	Ecstasy	Not suitable
23	G01, G06, G07, G10, G24	Pethidine	Marijuana	Not suitable
24	G03, G07, G09, G21	Morphine	Morphine	Suitable
25	G05, G12, G17, G22	Ecstasy	Ecstasy	Suitable
26	G07, G12, G15, G26	Heroin	Heroin	Suitable
27	G02, G08, G15, G23	Heroin	Heroin	Suitable
28	G06, G09, G15, G26	Morphine	Morphine	Suitable
29	G07, G11, G18, G23, G27	Codeine	Codeine	Suitable
30	G08, G15, G25, G26	Morphine	Morphine	Suitable

IV. CONCLUSION

The results of this study found that: (a) Machine learning in this study can predict drug users and types of drugs based on the symptoms that drug users complain about. (b) This study machine learning acquired knowledge about the symptoms of drug users, types of drugs, and basic knowledge related to the weight of the certainty factor of each type of drug and the symptoms caused so that it can diagnose drug users and the types of drugs used by users. (c) The accuracy of machine learning in this study in predicting the types of drugs used by users and the types of drugs used by users reached 80%. (d) The expert system in this research is website-based so that the expert system from this research can be used by various parties and in different places to identify users and the types of drugs used by users.

The implication of this research result is that the expert system built in this study can be a tool (choice) to replace or complete the testing system for drug users through urine testing in the laboratory.

The drawback of the results of this study is that machine learning expertise in this study is only limited to simple machine learning, as is the case with simple machine learning which was built on previous research by Zhongheng Zhang (2016), which used the KNN method in building learning machines. Furthermore, the machine learning expertise generated from this research is only limited to the expertise possessed in accordance with the knowledge obtained (symptoms, types of drug abuse, rule base, and calculation of certainty factor) under study. Therefore, further research needs to build machine learning that can increase its expertise based on more new data and use another method.

REFERENCES

[1] R. Jiménez, J. Anupol, B. Cajal, and E. Gervilla, "Data mining techniques for drug use research," Addict. Behav. Reports, vol. 8, pp. 128–135, 2018.

[2] F. M. Vassoler, E. M. Byrnes, and R. C. Pierce, "The impact of exposure to addictive drugs on future generations: Physiological and behavioral effects," Neuropharmacology, vol. 76, no. PART B, pp. 269–275, 2014.

[3] P. K. Shanmugam, "The Influence of Social Factors in Drug Addiction–A Mini Review of Work by Miller & Carroll (2006)," J. Alcohol. Drug Depend., vol. 05, no. 04, pp. 4–6, 2017.

[4] A. Boys, J. Marsden, and J. Strang, "Understanding reasons for drug use amongst young people: A functional perspective," Health Educ. Res., vol. 16, no. 4, pp. 457–469, 2001.

[5] G. López, L. M. Orchowski, M. K. Reddy, J. Nargiso, and J. E. Johnson, "A review of research-supported group treatments for drug use disorders," BMC Public Health, vol. 16, no. 51, pp. 1–21, 2021.

[6] Z. Justinova, L. V. Panlilio, and S. R. Goldberg, "Drug Addiction," Natl. Libr. Mediciine, vol. 1, no. 1, pp. 310–335, 2009.

[7] T. Saah, "The evolutionary origins and significance of drug addiction," Harm Reduct. J., vol. 2, pp. 1–7, 2005.

[8] G. Leshner, E. M.Stevens, S. Kim, N. Kim, T. L.Wagener, and A. C. Villantie, "Cognitive and affective responses to marijuana prevention and educational messaging," Drug Alcohol Depend., vol. 225, no. August, pp. 1–3, 2021.

[9] M. Zaman et al., "Drug abuse among the students," Pakistan J. Pharm. Res., vol. 1, no. 1, p. 41, 2015.

[10] L. W. Chou, K. M. Chang, and I. Puspitasari, "Drug Abuse Research Trend Investigation with Text Mining," Comput. Math. Methods Med., vol. 2020, pp. 1–8, 2020.

[11] A. Bevan and N. Patel, "An Electronic Prescription Alerting System-Improving the Discharge Medicines Process," Arch. Dis. Child., vol. 101, no. 9, p. e2.55-e2, 2016.

[12] A. Tsyben, N. Gooding, and W. Kelsall, "Assessing the Impact of a Newly Introduced Electronic Prescribing System Across a Paediatric Department – Lessons Learned," Arch. Dis. Child., vol. 101, no. 9, p. e2.13-e2, 2016.

[13] M. Mahmoudi, S. Pakpour, and G. Perry, "Drug-Abuse Nanotechnology: Opportunities and Challenges," ACS Chem. Neurosci., vol. 9, no. 10, pp. 2288–2298, 2018.

[14] J. Redman, "Recognizing the Warning Signs of Drug Addiction : What You Need to Know," Mental Health and Counseling Studies. pp. 1–7, 2021.

[15] N. Jojen, "A Survey Paper on Data Mining Techniques in Drug Industry," Int. J. Eng. Res. Technol., vol. 3, no. 30, pp. 296–299, 2015.

[16] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or organ category," Front. Chem., vol. 6, no. May, pp. 1–11, 2018.

[17] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," Ann. Transl. Med., vol. 4, no. 11, pp. 1–7, 2016.

[18] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," ICIC Express Lett., vol. 15, no. 4, pp. 319–326, 2021.

[19] A. Anggrawan, K. Hidjah, and Q. S. Jihadil, "Kidney failure diagnosis based on case-based reasoning (CBR) method and statistical analysis," in 2016 International Conference on Informatics and Computing, ICIC 2016, 2017, pp. 298–303.

[20] K. Muludi, R. Suharjo, A. Syarif, and F. Ramadhani, "Implementation of forward chaining and certainty factor method on android-based expert system of tomato diseases identification," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 9, pp. 451–456, 2018.

[21] C. P. C. Munaiseche, D. R. Kaparang, and P. T. D. Rompas, "An Expert System for Diagnosing Eye Diseases using Forward Chaining Method," in IOP Conference Series: Materials Science and Engineering, 2018, vol. 306, no. 1, pp. 1–8.

[22] N. Von Greiff and L. Skogens, "Recovery and identity: a five-year follow-up of persons treated in 12-step-related programs," Drugs Educ. Prev. Policy, pp. 1–10, 2021.

[23] A. Anggrawan, "Percentage of Effect of Blended Learning Madel on Learning Outcome," in Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019, 2019.

[24] A. Anggrawan, N. Ibrahim, S. Muslim, and C. Satria, "Interaction between Learning Style and Gender in Mixed Learning with 40 % Face-to-face Learning and 60 % Online Learning," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 5, pp. 407–413, 2019.