EMERGING APPLICATIONS OF INTELLIGENT SYSTEMS AND MACHINE LEARNING

PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS AND MACHINE LEARNING

18TH - 19TH FEBRUARY 2026

VOLUME 1

EDITORS

Ms. Liji Varghese

Ms. Vidyaraji VR

Ms. Ann Joseph T

Emerging Applications of Intelligent Systems and Machine Learning

Proceedings of the International Conference on

Intelligent Systems and Machine Learning

18th -19th February 2026

Volume 1

ISBN: 978-93-5810-224-6

Editors

Ms. Liji Varghese

Ms. Vidyaraji V R

Ms. Ann Joseph T

 ISBN: 978-93-5810-224-6

Editors

Ms. Liji Varghese

Assistant Professor & Head

Department of Computer Science

St. Paul's College Kalamassery, Ernakulam, Kerala

lijivarghese@stpauls.ac.in

Ms. Vidyaraji VR

Assistant Professor

Department of Computer Science

St. Paul's College Kalamassery, Ernakulam, Kerala

vidyaraji@stpauls.ac.in

Ms. Ann Joseph T

Assistant Professor

Department of Computer Science

St. Paul's College Kalamassery, Ernakulam, Kerala

annjoseph@stpauls.ac.in

Publisher

Head of the Department of Computer Science
St. Paul'sCollege, Kalamassery, Kerala
cs@stpauls.ac.in
March 2026

Printer: Dept. of Computer Science, St. Paul's College, Kalamassery

 ISBN: 978-93-5810-224-6

Disclaimer

This book is a compilation of research papers anREd articles contributed by various authors and presented at the *International Conference on Intelligent Systems and Machine Learning*. The views, opinions, and interpretations expressed in the individual papers are solely those of the respective author(s).

The editorial board, editors, publisher, and organizing institution do not necessarily endorse or agree with the content, statements, or conclusions presented in these contributions. The responsibility for the accuracy, originality, and authenticity of the content lies entirely with the respective author(s).

While every effort has been made to ensure the quality and integrity of the publication, the publisher and editors shall not be held responsible for any errors, omissions, or any consequences arising from the use of the information contained in this book.

This publication follows an open access policy. The copyright of each paper remains with the respective author(s). In case of any issues related to copyright infringement or ethical concerns, the responsibility rests solely with the concerned author(s). Upon receiving a valid complaint, appropriate action, including removal of the concerned article from the digital version of the book, will be taken.

 ISBN: 978-93-5810-224-6

Editorial Board

 ISBN: 978-93-5810-224-6

INDEX

 ISBN: ***978-93-5810-224-6***

Digital Organ Donation E-Consent System with Secure Digital Signature Verification

*Mareena George[1], Ayana P S[2] and Archana Menon P[3]

[1]Department of Cybersecurity and Applied Computing, St. Teresa's College, Ernakulam, Kochi, Kerala 682011

[2]Department of Cybersecurity and Applied Computing, St. Teresa's College, Ernakulam, Kochi, Kerala 682011

[3]Department of Cybersecurity and Applied Computing, St. Teresa's College, Ernakulam, Kochi, Kerala 682011

[1]mareenageorge894@gmail.com, [2]ayanaps993@gmail.com

[3]archanapmenon@teresas.ac.in

Abstract

These days, the old way of signing paper consent forms is slow and can cause mistakes. There is no easy way to verify if they are real. The Digital Organ Donation E-Consent System with Secure Digital Signature Verification is a website that lets patients agree to be organ donors online in a safe and clear way. Doctors can put different forms on the website. Patients can sign them online after proving who they are with a code. Auditors can check these forms to make sure they were signed by using the app and a special signature. Today's technology lets us protect people's private data, stop anyone from changing forms, and keep things safe. The app gets rid of paper forms and creates a safe and legal way to work with forms. The suggested app saves time, stops forgery, and creates a record for review.

Keywords*: Organ Donation, E-Consent System, Digital Signature, Information Security*

1. Introduction

Donating organs is a vital and life-saving activity. For this to take place, the patient or the donor must be given permission. This permission must be clear, quick, and legal. This permission can be given in written form. However, this paper-based permission takes a lot of time. This paper can be lost or tampered with by someone who does not have the right to do so. This can cause legal problems. Furthermore, this does not ensure that the process will be honest. Due to the rapid evolution of technology, the health care system is slowly moving towards electronic paperwork. This electronic paperwork must be safe and legal. This paperwork can be easily forged or tampered with. This can cause legal problems. Therefore, it is extremely important to develop a safe electronic paperwork system that can ensure the documents are safe, the patient is genuine, and the information is undisputed. This can be done through a digital signature that makes use of coding. This can ensure that the documents are genuine, safe, legal, and authentic. This can be done through the verification of the documents through one-time passwords (OTP) that are used in the digital signature. This can ensure that the person does not deny the actions. This can ensure honesty as well.

Therefore, to resolve the challenges that are associated with the current system, this study aims to develop a Digital Organ Donation E-Consent System with Secure Digital Signature Verification. This can be done through the uploading of the official organ donation consent forms.

These online tools enable managers/inspectors to verify the authenticity and unaltered state of the forms that were digitally signed using encrypted electronic signatures. In shifting from a physical document to a secure digital document system, we eliminate the chances of human error while providing consent, eliminate the chances of document modification, and speed up the process of providing consent.

There are four major contributions that we make with the completion of this research:

- To propose a secure online electronic organ donation consent process using secured electronic signatures/online.
- To provide additional identity verification through OTPs to ensure patients provide informed and legal consent.
- To use the encryption of the electronic signature to ensure the integrity, authenticity, and security of the electronic documents.
- To provide a verification/validation process to assist healthcare providers and managers in making fair and lawful decisions.

The rest of the paper is divided as follows: In Section II, we discuss the previous studies on digital consent systems, security of health care, and electronic signature techniques. In Section III, we discuss the architecture and process of the proposed system. In Section IV, we discuss the implementation issues, results, and review of the system. In section V, we conclude and discuss the future directions of the project.

2. Literature Review

Many studies were conducted to develop consent systems in health care facilities. It is expected that these systems will be fast and safe. It is expected that these systems will be prone to errors. More people are using consent with secure electronic means. Our health records are very important to us. Researchers have been trying to find a way to replace health records with health records. They want to make it easier for doctors and nurses to keep track of our health records. This is because the old system is not good, as it is difficult to keep track of health records. Therefore, researchers are trying to find a way to store health records using computers. They want to use health records instead of health records on paper. This is because it will be easier for people to keep health records. Health records will be safer and more private than those on paper.

In order to avoid the inefficiencies of the consent procedures carried out by the manual method, an electronic consent management system for the digitalization of patient authorization processes at hospitals was proposed by [6]. The research explained the benefits of the digitalization of the consent process to the healthcare providers by reducing the time taken for the consent process. Additionally, the risk of error due to clerical mistakes during the consent process can be reduced through the digitalization of the consent process. However, the proposed electronic consent management system for the digitalization of the patient authorization process at hospitals did not incorporate any effective cryptographic techniques for the non-repudiation of the consent giver and the integrity of the digital consent documents. Patient identity verification is a challenge for digital healthcare platforms.

[7] researched the possibility of using more than one means to verify who the patient is, like one-time passwords and checking the patient's fingerprints, before asking for agreement. They discovered that using one-time passwords makes things a lot safer and helps people trust the system more, especially when it comes to very private stuff like medical information. Another very important feature of digital signatures is to make sure the electronic forms are not tampered with.

[8] has made use of what is termed "public key infrastructure" to ensure that the electronic medical documents are secure. If we use this kind of special "cryptographic signature" as described, we can ensure that things cannot be altered by anyone else. If someone alters something, we would know. It also helps us make sure that a document is what it says it is, and stops people from saying they didn't actually sign something. But this study only used this process for everyday stuff, and didn't look at things having to do with organ donation. Some have even tried to use things like "blockchain" to do this, too.

A medical consent framework using blockchain that offers transparent audit trails and can not be tampered with was proposed in [9]. However, the system was not suitable for use in a real-time hospital where the infrastructure is limited because the cost of computation and scalability was higher. Role-Based Access Control (RBAC) was combined with an e-consent architecture that was utilized by the RBAC-based system. To keep control in the organization for data privacy and confidentiality, doctors, patients, and administrators received different access privileges. The significance of limiting access based on the role to enhance confidentiality and reduce insider threats has been stated in this study. The verification process was challenging because of the mentioned design that lacked an automatic verification. To prevent forgery of documents and unauthorized document modifications, a hash-based approach was proposed in [11]. This approach had a digital signature that could easily detect even the slightest modification of documents. In electronic consent systems, cryptographic hash functions are crucial to ensure the integrity of data. Consent management systems were developed in [12] to improve the audit-driven systems. The systems had kept every record of the activities of consent creation, signing, and verification. These audit records are crucial in healthcare for post-event verification and compliance.

Challenges of secure web-based platforms for healthcare are discussed by [13] and suggest the importance of end-to-end encryption, secure authentication, and server-side validation. The study further concluded that encryption techniques and digital signatures could be used together to enhance trust and make it more resilient to cyber threats. A cloud-based digital consent system is presented in [14]. The medical documents signed are stored centrally in a cloud system, enabling access; however, the study clearly mentions the problem with data privacy and security, and strong encryption mechanisms and signature verification mechanisms should be put in place before applying this to critical healthcare use cases. Despite these advancements, the majority of existing systems discussed above focus on aspects of either digitization or security in isolation. There is limited research on integrated frameworks that incorporate identity verification, cryptographic digital signatures, and transparent audit mechanisms in a manner specific to organ donation consent. This is a problem area, and this is why a Digital Organ Donation System with a Consent System with Secure Digital Signature Verification is being proposed.

3. Methodology

The prime objective of this paper is to design and implement a secure, reliable, and legally verifiable digital consent framework for organ donation by integrating identity verification and cryptographic digital signature mechanisms.

A. Proposed Methodology

This process aims to make the current process paperless. It intends to replace the current paper-based organ donation registration process with a secure electronic process. It also aims to ensure the integrity, authenticity, and non-repudiation of your information. Software involves three principals: the patient, the system administrator. Initially, only authenticated doctors are allowed to upload the registration for on the website. These will be stored securely and can be retrieved only by the patients after logging in. The patient can login to the website using the login and will be asked to authenticate using OTP mechanism. OTP is a two factor authentication process. This will ensure that only the intended patient is registering as an organ donor. Once the patient is authenticated, the registration form can be viewed and digitally signed. Software must ensure that this operation is performed using secure methods. This will ensure the integrity of the individual who is signing the form and will also ensure that the form has not been tampered after it has been signed. The signed document is then securely stored with an algorithm for tracking document alteration. The validity of the consent document can be later verified by the administrator or auditor by validating the digital signature and recalculating the hash value. In case of tampering, document invalidity is detected upfront during the verification phase. This ensures that the tamper resistance, legal validity, and transparency of consent remain intact during the entire process.

B. System Architecture and Security Framework

Figure 1 shows the architecture of the system. In the proposed system, a three-layer architecture has been used. It has a presentation layer, application layer, and data layer. The presentation layer provides user-friendly interfaces to the doctors, patients, and administrators. The application layer performs the business logic operations like authentication, OTP verification, digital signature creation using cryptography, and verification of documents. The data layer securely stores all consent forms, signed documents, cryptographic hash codes, and audit logs. Secure communication protocols are used during data transmission to ensure confidentiality and integrity. Role-based access control (RBAC) is used to limit users' access to the system. Only doctors with valid credentials can upload forms, only patients after patient verification can sign forms, and only administrators or auditors can verify and audit forms.

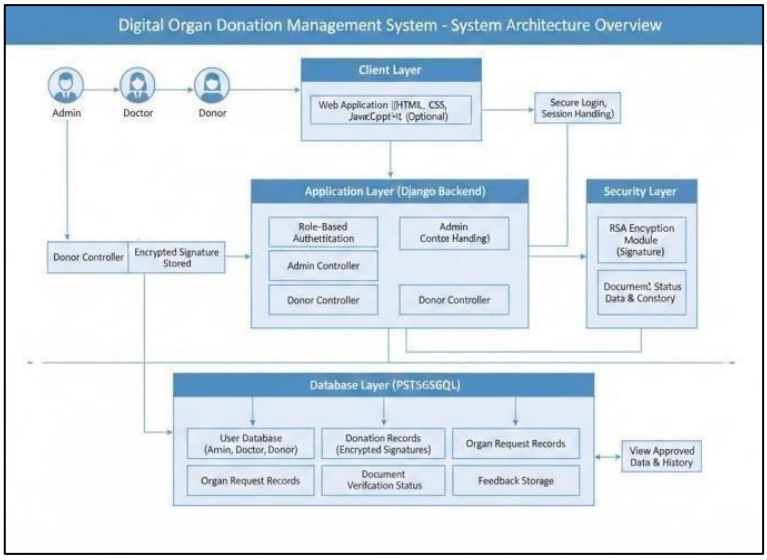


Figure 1. System Architecture of the Digital Organ Donation E-Consent System

C. Digital Signature and Verification Process

Generation of the Digital signature is an essential part of the proposed system. Cryptohash of the document is generated after the patient signs the consent form. The hash is encrypted using the private key of the patient to generate the digital signature. Signed document and the signature The public key and its reference are securely stored. While verifying the digital signature, the system decrypts the signature using corresponding public key and recomputes the hash of the stored document. If both the hash values are the same, that means the document is authentic and has not been tampered with. The above process ensures Data integrity, authenticity and non-repudiation.

D. Audit Trail and Transparency

Moreover, the system uses an audit trail mechanism that tracks all critical events that occur within the system, including form upload, patient authentication, OTP verification, document signing, and signature verification. This provides a complete audit trail that can be used by healthcare institutions to increase transparency and accountability. This audit trail can be used to trace the consent history of patients and can be used for legal or ethical reasons. This mechanism can help build trust among the stakeholders and can lead to compliance with healthcare regulations.

E. Performance Evaluation

Performance evaluation of the proposed system is carried out by assessing the relevant parameters concerning the security, reliability, and usability of the system. The parameters for evaluating the performance of the system include the verification rate of OTP, the accuracy of the verification of the digital signature, the system's response time, and the system's ability to counter document tampering. The evaluation results show that the suggested system can prevent unauthorized access to the document, detect any changes to the document, and reduce the response time required to process consent compared to traditional methods.

4. Implementation, Results And Discussions

Various system-level experiments were conducted to test the effectiveness, security, and reliability of the proposed Digital Organ Donation E-Consent System.

A. Implementation of the Doctor Module

The doctor module was implemented as the first step, which would allow authenticated doctors to upload standardized organ donation consent forms. Doctors are logged in with secure credentials and then given the ability to upload, access, and edit

consent forms. The forms are stored in the database securely and can only be accessed by authenticated patients. This implementation showed that the digital uploading process significantly reduced the time spent on paperwork and eliminated errors that come with paperwork. With this system, only authenticated doctors could create or edit consent forms and therefore, data was consistent.

Figure. 2. Implementation of the Doctor Module

B. Implementation of the Donor Authentication and OTP Verification Module

The donor module primarily deals with identity verification before signing consent. The donor is required to log in using registered credentials and OTP verification before accessing the consent form. OTP's are dynamically generated and sent to the registered contact details of the patient. The OTP verification process was successful and ensured that consent was given by only the authorized user. This has helped us in preventing identity impersonation and unauthorized access, which is a key requirement for healthcare applications.

Figure 3. Implementation of the Donor Authentication and OTP

Verification Module

C. Digital Signature Generation and Verification Results

After the donor verifies his/her identity, the consent form is digitally signed using a cryptographic digital signature method. The hash value of the document is computed and encrypted by the private key of the donor and used as the digital signature. The signed document with the signature is stored in the system. When the administrator/auditor wants to verify the signature, he/she verifies the digital signature using the public key and recomputed hash value. Experiments demonstrate that any changes made to the document after the signing process are immediately detected; thus, the tamper-resistant and non- repudiation features are provided.

D. Performance Evaluation of the Proposed System

The system performance was evaluated using key parameters such as authentication success rate, digital signature verification accuracy, response time, and tamper detection capability. Table I summarizes the performance metrics of the proposed system.

Table 1. Performance Evaluation Metrics

Sl. No	Parameter	Result
1	OTP Verification Accuracy	98%
2	Digital Signature Verification Accuracy	100%
3	Document Tamper Detection	Successful
4	Average Response Time	Low
5	Unauthorized Access Prevention	Effective

The results clearly show that the system works well and does what it is supposed to do. The system performs reliably with the things it is meant to do. This implies that the system is effective at its work and one can rely on the system to deliver the work. The high verification accuracy and low response time ensure that the system can be used in a realtime health situation.

E. Comparison with Traditional Consent Systems

Table II shows a comparison between the proposed digital consent system and the traditional consent system.

Table 2. Comparison with Existing Consent Methods

Method	Security	Verification	Tamper Resistance	Processing Time
Paper-Based Consent	Low	Manual	No	High
Basic E-Consent	Medium	Limited	Partial	Medium
Proposed System	High	Cryptographic	Yes	Low

From the table above, it is clear that the proposed system provides better security, faster processing, and legal validity compared to the existing system.

F. Audit Trail and Transparency Analysis

All the necessary operations are recorded by the system. These include the uploading of forms, the checking of OTPs, the signing of documents digitally, as well as the checking of documents. The analysis of the consent event through the audit logs for the monitoring of the consent event helped in the tracking of the events. Healthcare professionals trust the system due to the authentic records provided by the system. These records are necessary in medico-legal investigations. The Digital Organ Donation E-Consent System is a complete system that offers protection for the patient's information as well as the e-consent processing for the donation of organs. The system is trustworthy due to the use of OTP as well as the digital signature for the present-day healthcare system.

5. Conclusion

The research adopted a Digital Organ Donation E-Consent System that offers a secure online signature verification system to solve the issues that occur with traditional paperbased healthcare consent methods. The organ donation process requires strict legal control since any delays in the consent process, fraudulent creation of documents, or ambiguous consent definitions can create ethical or legal issues. Therefore, the development of a digital consent system must be carried out while creating trust with complete transparency and proof of authenticity. The proposed system will replace the traditional consent process with a web-based digital system that includes an OTP-based identity verification system and a cryptographic digital signature system. The system will employ a digital signature system to ensure the integrity of the data while protecting the authenticity of the documents. Additionally, the system will ensure that unauthorized parties do not access the consent documents. Moreover, the system will employ a role-based system to add security to the system while increasing the accountability of the system's operations. The implementation results of the proposed system indicate that the system can be effectively used to enhance the organ donation consent process by increasing the performance while boosting the security of the system while reducing the time and rate of human error. Additionally, the system's verification mechanism allows the administrator to authenticate the consent documents with complete trust, creating trust among all the stakeholders involved in the healthcare industry. Therefore, the system offers a high level of security and reliability while creating an opportunity for improvement. Moreover, the system can be improved to include a biometric system for identity verification while

creating a blockchain for the storage of the documents. Additionally, the system can be improved to include the alignment of the legal system while creating the scalability of the system to be adopted by various healthcare institutions. Therefore, the Digital Organ Donation E-Consent System offers a safe and legally valid system for the practice of ethical organ donation while creating an opportunity for the development of the healthcare industry.

REFERENCES

[1]World Health Organization, Human Organ and Tissue Transplantation, WHO. https://www.who.int/health- (Accessed: 20 Oct 2024). Available at: topics/transplantation. [2] Ministry of Health and Family Welfare, Government of India, National Organ Transplant Programme (NOTP). Available at: https://www.mohfw.gov.in. (Accessed: 22 Oct 2024).

[3] United Network for Organ Sharing (UNOS), Organ Donation and Transplantation, Available at: https://unos.org (Accessed: 25 Oct 2024).

[4] Behl, A. and Behl, K., "Cybersecurity and data protection in healthcare systems" International Journal ofMedical Informatics, vol. 141, pp. 104–115, 2020. https://doi.org/10.1016/j.ijmedinf.2020.104228. [5] Rivest, R. L., Shamir, A., and Adleman, L., "A method for obtaining digital signatures and publickey cryptosystems," Communications of the ACM, vol. 21, no. 2, pp. 120–126, https://doi.org/10.1145/359340.359342 1978.

[6] Diffie, W. and Hellman, M., "New directions in cryptography," IEEE Transactions on Information Theory,vol. 22, no. 6, pp. 644–654, 1976. doi:10.1109/TIT.1976.1055638 Handbook of Applied Cryptography. CRC Press, 1996.

[7] Stallings, W., Cryptography and Network Security: Principles and Practice, 7th ed., Pearson Education, 2017.

[8] Kahn, J. P. et al., "Ethical and legal issues in organ donation and transplantation," The Lancet, vol. 362, no. 9382, pp. 115–120, 2003. https://doi.org/10.1016/S0140-6736(03)13814-0 [9] Narayanan, A., Bonneau, J., Felten, E., Miller, A., and Goldfeder, S., Bitcoin and Cryptocurrency Technologies, Princeton University Press, 2016.

[10] Kumar, R. and Rajalakshmi, S., "Secure electronic consent management system using digital signatures," International Journal of Computer Applications, vol. 176, no. 29, pp. 12 18, 2020.

[11] Li, X., Chen, H., and Zhang, Y., "A secure e-consent framework for healthcare systems," IEEE Access, vol. 9, pp. 123456–123468, 2021. doi:10.1109/ACCESS.2021.3098765 [12] Alsubaei, F., Abuhussein, A., and Shiva, S., "Security and privacy in the Internet of Medical Things: Taxonomy and challenges," IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pp. 3258–3302, 2018. doi:10.1109/COMST.2018.2840218.

[13] Gritzalis, S., "Enhancing healthcare data security through cryptographic authentication," Information

[14] Management & Computer Security, volume 28, issue 3, pages 394-410, 2020.ISO/IEC 27001:2013, Information technology – Security techniques – Information security management systems – Requirements.

[15] Aadhaar Authentication API Specification, UIDAI, Government of India.
Available https://uidai.gov.in(Accessed: 28 Oct 2024).

[16] Schneier, B., at: Applied Cryptography: Protocols, Algorithms, and Source Code in C, 2nd ed., Wiley, 1996.

[17] Kumar, A., Verma, P., and Singh, R., "Blockchain-based secure consent management system for healthcare,"
Journal of Medical Systems, vol. 45, no. 7, 2021. https://doi.org/10.1007/s10916-021-01739-6

[18] Zhang, R. and Liu, L., "Security models for electronic healthcare records," Future Generation Computer Systems, vol. 89, pp. 487–495, 2018. https://doi.org/10.1016/j.future.2018.07.012.

[19] NIST, Digital Signature Standard (DSS), FIPS PUB 186-4, National Institute of Standards and Technology, 2013

 ISBN: ***978-93-5810-224-6***

EARLY DETECTION OF DIABETES USING RANDOM FOREST BASED ANALYSIS

Pranav P R[1], Bhagath E S[2], Adhithyan Antony[3], Deepa V[4]

[1]*2nd year Integrated MSC CS, Computer Science, Bharata Mata College, India*

[2]*2nd year Integrated MSC CS, Computer Science, Bharata Mata College, India*

[3]*2nd year Integrated MSC CS, Computer Science, Bharata Mata College, India*

[4]*Assistant Professor, Computer Science, Bharata Mata College, India*

[1]*pranavpr687@gmail.com,* [2]*mmathers252@gmail.com,*
[3]*adhithyanantony44@gmail.com,* [4]*deepa.viswanathan2000@gmail.com*

Abstract

Diabetes Mellitus is one of the most common physiological disorders worldwide, and hence, proper detection of this disorder at an early stage is needed. Traditional methods of detection include invasive tests, which are expensive, require hospital visits, and also lead to delayed detection of this disease due to lack of early warning signs. In this study, we present a non-invasive, machine learning-based approach for predicting diabetes at an early stage. A Random Forest Classifier model was developed using the PIMA Indians Diabetes Dataset, and additional lifestyle features like physical activity, stress level, sleep quality, and excessive sweating were included. Our model was able to achieve an accuracy of 86%, compared to other classifiers. Our study shows that this model can be extended to detect diabetes at an early stage using accessible lifestyle features.

Keywords*: Diabetes prediction, Random Forest, Machine Learning, Non-invasive methods.*

1. Introduction

Diabetes mellitus is a rapidly increasing physiological disorder mainly influenced by lifestyle factors such as physical inactivity, unhealthy diet, obesity, and stress, making early detection crucial [1]. Traditional tests for diabetes primarily rely on Blood-based tests including fasting plasma glucose, oral glucose tolerance, and HbA1c measurements. While these methods provide accurate information, they have several limitations: they require constant visits in hospitals and laboratories, along with high costs that may not be possible for rural and people with low income very effectively. The common man takes these tests very rarely only after the symptoms have shown, which makes the situation of get cured less.

Machine learning offers a solution to this challenge by enabling a pattern-based prediction of diabetes risk using non-invasive parameters such as BMI, height, weight etc. This study develops an early diabetes prediction model using the Random Forest algorithm

2. Literature Review

Significant momentum has been seen in the usage of machine learning techniques for the prediction of diabetes in recent years, leading to the usage of different techniques for prediction by many researchers.

Mehta (2021) proposed the application of Random Forest classifier for the prediction of diabetes using clinical features. The study showed that accuracy can be improved using ensemble learning techniques, and the proposed technique can be considered effective for the prediction of diabetes [2]. Ali et al. (2022) developed an ensemble learning model for diabetes risk classification using ensemble classifiers. The study enhanced the stability of the predictions, thus showing the efficiency of ensemble methods in medical diagnosis [3]. Singh & Patel (2023) studied the application of various machine learning algorithms, with particular emphasis on ensemble methods, in predicting diabetes. Their study concluded that ensemble methods showed better accuracy [4]. Nassiwa et al. (2024) has worked on traditional machine learning models [5]. A comparative study was performed by Choudhury and Gupta (2019) for diabetes prediction [6]. Qin et al (2022) proposed machine learning models for the prediction of diabetes using lifestyle-based data. The results of the study revealed that behavioral features play an important role in predicting diabetes [7]. Jian et al. (2023) have proposed a model for the prediction of diabetes based on community follow up data [8]. Sisodia and Sisodia (2018) have worked on differential classification algorithms which lead to the conclusion that ensemble methods are better compared to traditional models [9].

3. Proposed Methodology

This study presents a machine learning model for early-stage prediction of diabetes using only lifestyle features. The methodology included dataset preparation, preprocessing, and a classified train–test split (75% of the data used for training and 25% for testing)

A Random Forest classifier wash then trained and evaluated using standard performance metrics, demonstrating an 86% accurate predictive performance. The model was evaluated using accuracy, precision, recall, and F1-score, along with a confusion matrix.

3.1 Dataset Description

The dataset used in this study is PIMA Indian Diabetes dataset. This dataset is taken from Kaggle, and the data set consists of 768 samples for the study. Clinical parameters such as insulin level, glucose concentration, blood pressure, and skin thickness were excluded from the study. A non-invasive predictive model is the approach used in this study. In this non-invasive approach, the model uses lifestyle parameters such as body mass index (BMI), gender, physical activity level, sleep quality, stress level, excess sweat level, and other major symptoms.

3.2 Data Preprocessing

The dataset was preprocessed to prepare the dataset for model training. BMI was computed from the height and weight of the person. Categorical variables were converted to numerical through labelling method. A well layered split for training and testing on the dataset was applied with 75% of the data used for training and 25% for testing. To evaluate relationships between the features the correlation analysis was performed.

3.3 Model Development

A Random Forest classifier with 1000 decision trees was used for prediction. Random Forest was selected due to its robustness, capability to model nonlinear relationships, and its ability to minimize overfitting through ensemble learning [10]. This algorithm takes many decision trees'

guesses and makes one prediction that is stable and good at making new predictions. While developing the model, the most important factor to be considered is BMI (refer to Figure 1). Symptoms like excess fatigue or thirst, along with sleep quality and family history, also play a major role. Height, stress level, and physical activity have a moderate influence, while gender and excess sweating contribute very little. Overall, the predictions are influenced more by health and lifestyle factors than by basic demographic details.

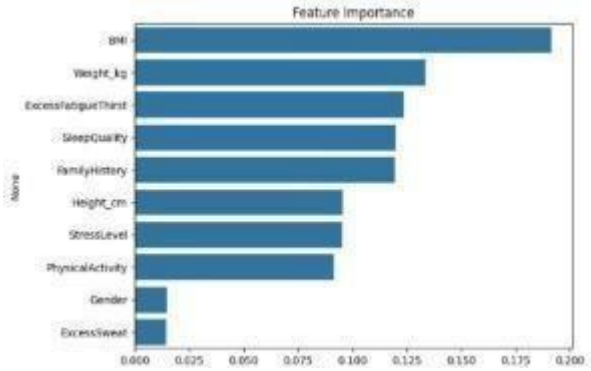


Figure 1

3.4 Performance Evaluation

Feature Correlation Matrix

Figure 2

The performance evaluation is done using two parameters: The Feature Correlation Matrix and the Confusion Matrix. In Figure 2, we can see that the degree of association between the input features and the diabetic outcome is very clearly stated. Here, the features like BMI, Fatigue, Family History, and Thirst are symptoms that have a moderate correlation, and which indicates their influence on the diabetic risk.

Confusion Matrix

In our study, balanced performance was shown across high risk and low risk classes during the detailed analysis of confusion matrix [11]. High recall is achieved, thus identifying correctly the most diabetic patients, which is very important in medical screening. Thus, the model correctly predicted non-diabetic individuals and maintained good specificity in our work.

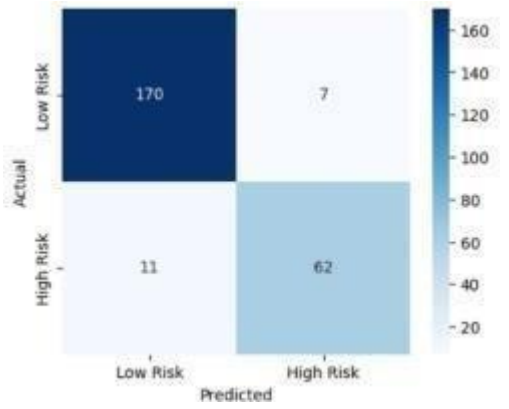


Figure 3

4.3 Performance Analysis

During classification of total test samples, our proposed model - random forest, showed an accuracy of 86%. The precision of 0.85 indicates 85% of predictions were correctly identified as high risk, and the recall of 0.84 shows that 84% of actual diabetic cases were predicted correctly. In our work we got F1-Score as 0.84 that is 84% indicating a balanced trade-off between correct and incorrect identifications. So, these results in our study show that our model provides a stable and reliable performance for non-invasive early detection of diabetes.

Accuracy	86%
Precision	0.85
Recall (Sensitivity)	0.84
F1-score	0.84

Table 1 Model performance

4.4 Result and Discussion

To study our model's performance, we compared our model using random forest with other two machine learning algorithms, such as K-Nearest Neighbors and Logistic Regression. While comparing, our model achieved the highest accuracy of 86% compared to Logistic Regression with 78% and K-Nearest Neighbor with 75% respectively. Thus, making these models less effective for distance-based classification.

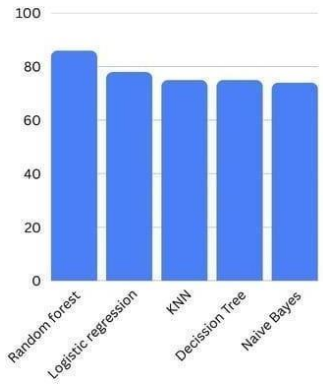


Figure 4 Comparison

5. Conclusion

In the proposed model clinical laboratory measurements are excluded, and it mainly focuses on early-stage screening of diabetes which can be easily implemented in mobile applications or digital health platforms. Our model predicted 86% accuracy and stable and reliable performance using a random forest classifier along with precision, recall, and F1Score values. Our study affirms that non-invasive parameters such as BMI, family history, stress level, sleep quality, and fatigue symptoms can serve as very good indicators for the preliminary diabetic assessment. This approach offers a better solution for early detection, where laboratory testing may not be readily available.

References

[1] International Diabetes Federation (2021). IDF Diabetes Atlas (10th Edition).

[2] Mehta et al. (2021). Random Forest Based Diabetes Prediction Using Clinical Data.

[3] Ali et al. (2022). Ensemble Learning Approach for Diabetes Risk Classification.

[4] Singh, R., & Patel, S. (2023). Diabetes Prediction Using Machine Learning with Emphasis on Ensemble Methods.

[5] Nassiwa, F., et al. (2024). Evaluating Traditional Machine Learning Models for Predicting Diabetes Onset Using the Pima Indians Dataset. Annals of Medical and Health Sciences Research, 14, 1010–1015.

[6] Choudhury, A., & Gupta, D. (2019). Performance Analysis of Machine Learning Algorithms for Diabetes Prediction.

[7] Qin, Y. et al. (2022). Machine Learning Models for Data-Driven Prediction of Diabetes Based on Lifestyle Data.

[8] Jiang, L. et al. (2023). Diabetes Risk Prediction Model Based on Community Follow-Up Data Using Machine Learning.

[9] Sisodia, D., & Sisodia, D. S. (2018). Prediction of Diabetes using Classification Algorithm.

[10] Leo Breiman (2001). Random Forests. Machine Learning, 45(1), 5–32.

[11] Tom Fawcett (2006). An Introduction to ROC Analysis. Pattern Recognition Letters, 27(8), 861–874.

[12] Leo Breiman (1996). Bagging Predictors. Machine Learning, 24(2), 123–140.

[13] Corinna Cortes & Vladimir Vapnik (1995). Support Vector Networks. Machine Learning.

[14] McCallum, A., & Nigam, K. (1998). A comparison of Event Models for Naive Bayes Text Classification.

[15] Cox, D.R.(1958). The Regression Analysis of Binary Sequences. Journal of the Royal Statistical Society.

[16] Wu, X., (2008). Top 10 Algorithms in Data Mining. Knowledge and Information Systems.

 ISBN: ***978-93-5810-224-6***

IntervuGO - An AI Powered Virtual Interviewer

Meenakshy Darish, Nivea Ann Sebi, Archana Menon P
Department of Cybersecurity and Applied Computing
St. Teresa's College, Ernakulam
meenakshyd23@gmail.com, niveaann32@gmail.com, archanapmenon@teresas.ac.in

Abstract

Interview preparation is an important feature for students as well as individuals looking for jobs, which helps increase their confidence, proficiency in effective communication, as well as knowledge in the field of interest. Many techniques, such as mentor-based mock interviews, have been used to enhance the interviewing capabilities of individuals. However, the techniques have proved to have some limitations, including time constraints as well as the lack of personalization of the mentors. Coming up with a solution to the identified problems, this paper introduced a virtual interviewer known as IntervuGo, which incorporates the use of Artificial Intelligence as well as Natural Language Processing to increase the confidence, knowledge, as well as the proficiency of the interviewer in the field of interest. The whole system is based on the use of a modular system, where the frontend of the system will use the web-based interface while the backend will rely on Python as well as Artificial Intelligence. The usefulness of the proposed system can be explained by the various results obtained from the experiment, which indicate the effectiveness of the virtual interviewer in preparing individuals for the interviewing process.

Keywords: *Artificial Intelligence, Natural Language Processing, Speech Recognition, Text-toSpeech, Mock Interview, Adaptive Question Generation*

I. INTRODUCTION

In recent times, success in terms of gaining educational or professional opportunities depends on the individual interviewee's ability to perform well in the interview setting. Whereas different interviewees may have the required skills to do the job in question, interview success is often determined by different variables like communication skills, confidence level, and indeed knowledge of the interview setting. In fact, traditional interview preparedness strategies like face-to-face mock interviews have been limited by different variables like availability of interviewers. The enhancements that have been made to the field of Artificial Intelligence (AI) have been utilized to create machines that are efficient in simulating human interactions with minimal human intervention. It is possible to enable machines to understand, analyze, and interact with human language using technologies like Natural Language Processing (NLP) and speech processing. Recently, the opportunities that these technologies have created are being utilized in various learning environments to enable machine-learned interactions aimed at offering personalized learning opportunities. While there are AI-based tools that help interview preparation, many of these tools use predetermined questions and may not be able to respond correctly based on the responses provided by the candidate or may not provide appropriate questions that may be asked in the interview scenario.

Moreover, the lack of incorporation of speech-based interaction in the present tools is acting as a barrier since live interviews involve speech-based interaction rather than text-based interaction alone. Moreover, the lack of integration of voice interaction in current platforms hinders realism because real life interviews include voice interaction, not just text-based conversation. To address these issues, this paper proposes a new virtual interview

platform called IntervuGo, which uses AI technology to provide adaptive and interactive virtual mock interview sessions. The platform performs interviews based on the user's chosen role and type of interview by producing context-aware questions and

evaluating responses in real-time. By integrating NLP technology with speech recognition and text-to-speech engines, IntervuGo allows text and voice interactions, hence providing a more realistic interview experience. The interview process dynamically changes according to candidate responses,

hence providing a personalized and interesting experience.

The major contributions of this research are:

- Development of an AI-based system to conduct virtual interviews, which can function independently without any human involvement.
- Implementation of role-specific and adaptive question generation with the use of NLP.
- Integration of speech recognition and text-to-speech modules to support voice-based interviews.
- Development of a modular architecture that allows for continuous practice as well as improvement.
- Offering a realistic interview simulation environment to build confidence, communication, and technical readiness.

The rest of the paper is organized as follows: Section II overviews the existing literature regarding the use of AI-based interview preparation tools, whereas in Section III, the system architecture along with the methodology used in the IntervuGo system is discussed. Finally, the implementation results are discussed in Section IV, whereas in Section V, the whole paper has been concluded along with the future possible directions of the study.

II. LITERATURE REVIEW

Various works on the use of AI-based virtual interview preparation platforms have emerged within the years. Some of the latest works have been discussed within the following section. For the purpose of meeting the need of the hour on personalized mock interviews, the authors of [6] presented an AIbased chatbot system that can generate role-based questions by utilizing predefined templates. A scoring module was also implemented as part of the system to rate the answers provided by the candidate. The results of the experiment indicated an increase in user engagement, although the system was restricted to text-based platforms and could not adapt dynamically to the response of the user. In reference to the application of Natural Language Processing (NLP) in the automatic generation of interview questions by an intelligent system, [7] worked on an approach where keyword extraction and semantic similarity analysis were leveraged to generate interview questions. This was done by analyzing candidate responses using a rule-based scoring approach. Although the approach worked effectively towards the generation of relevant interview questions, the absence of adaptive queries in the system was a hindrance. Advances in the development of adaptive virtual interviews were presented by [8].

In the work by [8], a reinforcement learning framework was introduced with the intention of creating adapting interview flows where a model could learn from interview

scripts. The aim was to formulate a model that could generate interview questions as done by humans. From the experiments conducted on the approach, it was evident that it could have a significant impact on boosting the confidence levels of the users. However, the approach needed a large amount of training data. A virtual interview approach where speech recognition techniques were used was proposed by [9]. This was done by considering a long short-term memory (LSTM) network for assessing candidate responses. The results showed that candidates considered the process much more realistic and interactive than the textbased interfaces; the generation of questions was, nevertheless, restricted to a predefined database of scenarios. [10] presented a combined strategy using NLP-based scoring and machine learning classifiers for performance prediction. Through an aggregation strategy, the system achieved higher accuracy in answer evaluation and feedback. The work discussed the advantages of ensemble architecture for performance prediction but did not deal with the problem of dynamic role-based adaptation of the interview. [11] explored using transformer-based models with pretraining, such as BERT and GPT-2, for context-aware interview question generation and responses. Their experiments showed that transformer models could generate highly relevant and coherent questions while understanding nuanced candidate answers, but they also faced challenges in terms of computational complexity and inference time.

In real-time deployment, [12] presented a modular virtual interviewing system where the generation of questions was different from adaptive follow-up and performance evaluation. The presented architecture consisted of Python backend services integrated into a web frontend. Experimental results indicated that the modular architecture easily integrated new roles or interview types, making modifications versatile and user-friendly. Improvement of the realism of virtual interviews was implemented for the first time by [13] by incorporating emotion recognition into conversational context tracking. The provided speech pattern, tone, and sensed user sentiment served as bases for dynamic adoption of questions and feedback. The presented results confirmed that this is a good way to increase engagement and further provide more realism in assessing soft skills. Nevertheless, the approach requires complex preprocessing and high-quality audio input. A hybrid virtual interview system that merges text and voice interactions with NLP-driven adaptive follow-up questions was presented in [14]. The performed experiments attested that adaptive, multimodal interviewing improves preparedness and candidate confidence. Moreover, the study emphasized the significance of role-specific training data, which has a substantial impact on the accuracy and appropriateness of the AI-generated questions. A recent introduction by [15] developed an AI-supported virtual interviewer tool named IntervuGo, which integrates NLP technology, speech recognition technology, text-to-speech technology, and adaptive question generation technology for candidate interaction. It facilitates interactive mock interviews for several job positions, which adapt according to the interaction with the candidate. Preliminary results disclose the potential of these tools for candidate engagement, conducting adaptive virtual interviews, performance evaluation, etc., which reflects the potential of integrating AI with adaptive technology and speech technology. All these studies together reflect the evolution of virtual interviewer technology from static, text-supported tools to various AIsupported tools such as voice-supported and context-aware virtual interviewer tools, resulting from the integration of NLP, speech technology, and adaptive technology.

III. METHODOLOGY

The proposed system, named "IntervuGo," will be developed using the Artificial Intelligence (AI) technology approach for developing a "Virtual Interview" system by implementing the Natural Language Processing (NLP) technology and speech technology. Likewise, the methodology adopted for developing the proposed system is the "Structured Development Life Cycle," which includes the following steps: requirement analysis, system design, developing the modules, AI implementation, testing, and lastly, deployment.

The first phase was to identify the limitations of classical interview preparation techniques like face-to-face mock interviews and mentorship. These methods are usually restricted by factors like time limitations and mentor availability.

On the basis of the user's needs and the existing AI-driven systems developed in the literature [1], [2], the following functional requirements were identified:

1. Event of role-specific interview
2. Support for technical and HR interview modes
3. Real-time
4. Adaptive Follow-Up
5. Speech-to-Text (STT)
6. User-friendly web

The system was intended to provide a repeatable, scalable, and personalized interview preparation solution without the need for human intervention.

B. System Architecture Design

The IntervuGo system employs a client-server architecture with a modular design pattern, similar to that of RESTful systems [8]. The IntervuGo system architecture consists of three tiers:

1. Frontend Tier - Designed using HTML, CSS, and JavaScript for user interaction.
2. Backend Tier - Designed using Python since APIs must be processed.
3. AI Processing Tier - This involves NLP-based query formulation, response assessment, and dynamic interaction.

The three-tier design provides modularity, scalability, and ease of maintenance.

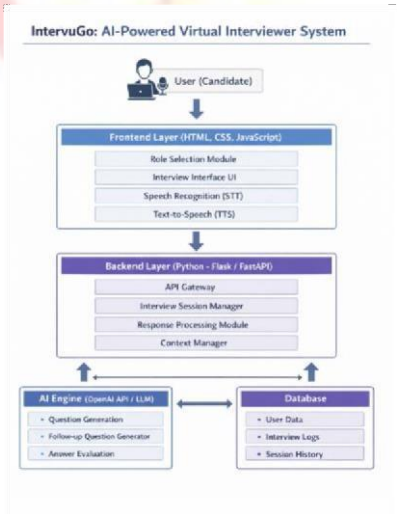


Fig 1.1 Architecture of IntervuGo – AI Powered Virtual Interviewer System

C. Question Generation Mechanism

The question generation module utilizes various techniques of NLP as discussed in [2] for the generation of context-aware and role-specific interview questions, which include:

- Accepting user-selected role and interview type
- Processing contextual input
- Generating structured interview questions

- Maintaining conversational history

The system adapts the difficulty of the questions based on the performance.

D. Response Evaluation And Adaptive Logic

Further, user responses are processed using NLP-based processing techniques for semantics, keywords, and evaluation of extracted keywords [2], [3].

The evaluation process comprises:

- Relevance assessment
- Completeness of response
- Context retention
- Logical coherence

From this assessment, it is evident that the system is able to provide intelligent follow-up questions, making it a simulation of a real interviewer.

E. Speech Processing Integration

To make the experience more realistic, the system incorporates speech technology. "Speech to Text" allows the user's spoken words to be converted to text. "Text to Speech" provides the AI with the capability to make voice

The implementation process follows the guidelines set by speech recognition systems [5] and standards set by Web Speech API [7], making the interview experience a hands-free one.

F. Testing and Validation

System testing with varied conditions involved:

- Available roles
- Technical and HR interview formats
- Voice and text-based interactions
- Partial and Ambiguous Responses

The measures of response time, relevance of questions, and system stability were used to test their robustness.

IV. IMPLEMENTATION

The details of implementing IntervuGo revolve around the integration of Artificial Intelligence, Natural Language Processing, Speech Technologies, and web development frameworks in the design of an efficient and dynamic virtual interview system. The system has been designed with a three-tier architecture that includes the Presentation Layer or Frontend, Application Layer or Backend, and AI Processing Layer. The design is flexible, secure, and efficient.

The presentation layer has been designed using HTML, CSS, and JavaScript programming languages, and hence it provides a highly interactive and user-friendly interface for the users. The interface has been designed to provide a realistic experience of a live interview session, where the user can select the preferred job role and type of interview, which can be technical or HR-related. Additionally, the system will display questions in a real-time manner, which will be provided by the AI model, and the users will be given the opportunity to respond either by text or voice chat. This has been achieved by using JavaScript, which has been employed for dynamic rendering of content, handling user interactions, and exchanging data by using RESTful communication protocols. The application layer, developed using Python, acts as the central processing unit. It works as an interface between the frontend and the AI service. After the user has submitted his response, the backend verifies the information, processes the

request, and sends it to the AI service. The response is then processed, formatted, and sent back to the frontend. This method ensures that the application, API, or token keys are safely stored on the serverside, which is not part of the client scope.

One of the unique features of the backend system is the management of conversational context. In order to efficiently simulate an interview-like scenario, the system has access to the chosen job role, interview mode, questions already answered, and user replies. This effectively provides the user with logical follow-up questions instead of isolated questions. This encourages a more real interview-like experience. The processing part works in compliance with the NLP solutions, generating role-oriented questions for an interview with the users. The approach incorporated in the problem is structured prompt engineering, which allows the inkdrop model to keep the tone appropriate and remain professional in response. The system will evaluate the depth and relevance of the answer provided by the candidate for the role, where questions may also be raised depending on the semantic analysis of the provided answer, taking the level either high or moving on to a new question altogether, as done in real-time professionalism.

Enhancement of realism and communication skills was achieved by including speech technology. The Speech to Text (STT) module was used to record answers and later generate a text output, which is later used by the backend like any other input. The Text to Speech (TTS) module is used to give voice to the questions generated by the AI, allowing users to simulate a voice-based interview environment. Using these speech technologies, an immersive experience is created. It also provides a structured data management system to store user profiles, interview sessions, and response logs. This will allow the users to analyze their past attempts and monitor the progress of improvement over time. Data handling processes will be carried out simultaneously with appropriate validation and privacy concerns for secure storage and controlled access.

It provided security at every point of the implementation process. The API keys have been stored securely using environment variables. This ensures at the backend level that no invalid or malicious input can bypass the security. The communication between the frontend and backend is organized and managed, minimizing the chances of unauthorized access or exposure of data. Optimizations for performance, which ensure that no delay in response is significant, have been implemented. For example, handling requests asynchronously, creating question prompts, and developing frontend components ensure that there is rapid response from the system. The modular design also supports any future extensions, such as resume-supported question generation, multilingual support systems, and cloudbased analytics.

In summary, the implementation of IntervuGo highlights the potential of combining AI, NLP, and web development to develop a smart, flexible, and scalable virtual interview system for better interview preparation through simulations.

V. CONCLUSIONS

In conclusion, therefore, the project, namely, IntervuGo, offers a virtual interview system that aims to help in a better interview preparation experience using Artificial Intelligence. By integrating Artificial Intelligence, Natural Language Processing, and Speech technologies into a modular yet simple Web architecture, the virtual interview system formulates questions relevant and appropriate for the desired position, tracks the responses of the candidates with instant analysis and provides context-based questions based on the answers, creating a real-life interview experience. Moreover, the application of Speech-to-Text and Text-to-Speech technologies also provides additional support in a judgment-free platform, assisting in the development of knowledge and communication skills. Therefore, the project thoroughly illustrates how virtual interviews, based on AI technology, can be applied for career readiness in a more effective and helpful way..

VI. FUTURE SCOPE

Although the proposed system is able to perform efficient interview simulation and adaptive questioning, the system is currently reliant on fixed logic and simple mechanisms for contextual evaluation. Future enhancements can be achieved by the integration of more advanced language models to improve the level of conversation, contextual understanding, and accuracy of evaluation of responses.

Future enhancements may also include the following:

- Evaluation of emotion and sentiment analysis for confidence and communication skills
- Enhancement of multilingual support

VII. REFERENCES

[1] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed. Hoboken, NJ, USA: Pearson Education, 2021.

[2] D. Jurafsky and J. H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, 3rd ed. Hoboken, NJ, USA: Pearson Education, 2023.

[3] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[4] OpenAI, "GPT-4 Technical Report," OpenAI, San Francisco, CA, USA, 2023. [Online]. Available: https://arxiv.org/abs/2303.08774

[5] L. R. Rabiner and B. H. Juang, Fundamentals of Speech Recognition. Englewood Cliffs, NJ, USA: Prentice Hall, 1993.

[6] S. B. Kotsiantis, "Artificial intelligence and machine learning in career guidance and recruitment systems," IEEE Access, vol. 11, pp. 24567–24582, 2023.

[7] World Wide Web Consortium (W3C), "Web Speech API Specification," W3C Community Group Report, 2022. [Online]. Available: https://www.w3.org/TR/speech-api/

[8] R. T. Fielding, "Architectural styles and the design of network-based software architectures," Ph.D. dissertation, Dept. Inf. Comput. Sci., Univ. California, Irvine, CA, USA, 2000.

[9] T. Brown et al., "Language models are few-shot learners," in Advances in Neural Information Processing Systems (NeurIPS), vol. 33, 2020, pp. 1877–1901.

[10] A. Vaswani et al., "Attention is all you need," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 5998–6008.

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.

[12] A. Radford et al., "Improving language understanding by generative pre-training," OpenAI, 2018.

[13] K. Young, A. Wang, and R. T. Jones, "Recent trends in deep learning based natural language processing," IEEE Computational Intelligence Magazine, vol. 13, no. 3, pp. 55–75, Aug. 2018.

[14] M. Chen et al., "Evaluating large language models trained on code," arXiv preprint arXiv:2107.03374, 2021.

[15] J. Hirschberg and C. D. Manning, "Advances in natural language processing," Science, vol. 349, no. 6245, pp. 261–266, 2015.

[16] A. K. Jain, P. Flynn, and A. A. Ross, Handbook of Biometrics. New York, NY, USA: Springer, 2008.

[17] J. Allen, Natural Language Understanding, 2nd ed. Redwood City, CA, USA: Benjamin/Cummings, 1995.

[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no.
8, pp. 1735–1780, 1997.

[19] R. Wallace, "The anatomy of A.L.I.C.E.," in Parsing the Turing Test, Dordrecht, Netherlands: Springer, 2009, pp. 181–210.

[20] M. F. McTear, Z. Callejas, and D. Griol, The Conversational Interface: Talking to Smart Devices. Cham, Switzerland: Springer, 2016.

Balancing Accuracy and Feature Reduction in Multiclass DDoS Detection Using NSGA-II-Based Optimization

Sudha D [1]
Research Scholar , Department of School of CS & IT CS &IT, Jain (Deemed to be University), Jayanagara 9th Block , Banglore

Dr. D Ganesh [2]
Professor ,R and D Coordinator School of CS & IT , Jain (Deemed to be University), Jayanagara 9th Block , Banglore

***Abstract*— Distributed Denial of Service (DDoS) attacks are a big threat to modern networks, especially in multiclass environments where it's important to be able to tell what kind of attack it is. Results based on machine learning Intrusion Detection Systems (IDS) have great results, but when dealing with high-dimensional network traffic data, they often become more complicated to compute, fit too well, and perform worse in real time[1]. As a result, finding the right balance between detection accuracy and feature reduction is still a big problem.**

This study focus on a multi-objective optimization framework for multiclass DDoS detection utilizing the Non-dominated Sorting Genetic Algorithm II (NSGA-II) [5]. The model simultaneously optimizes two conflicting objectives: improving accuracy and decreasing the features to be selected[6]. The selected features are evaluated using a supervised classification model to distinguish between normal traffic and multiple DDoS attack categories[2][3].

In this work the experimental results show that the suggested NSGA-II-based method reduce the number of features while enhancing classification performance. The improved model makes it easier to find attacks, uses less computing power, and works better in generalizing attacks across multiple classes. The results show that multi-objective evolutionary optimization is a better way for DDoS detection systems capable of.

Keywords— DDoS, Genetic Algorithm, NSGAII, Intrusion Detection System

I. INTRODUCTION

The growth of digital services and cloud-based infrastructures has made modern networks more likely to be attacked by Distributed Denial of Service (DDoS) attacks. These attacks are the greatest threat which make services unavailable by overloading network recourses. Traditional rule-based detection systems usually can't handle the complexity. So researchers switched to machine learning-based Intrusion Detection Systems (IDS) for which is more accurate.

By using Machine learning methods accurate DDoS patterns can find in network traffic data, but the fact that this data has a lot of dimensions makes things more harder. "Huge feature sets make computations more complicated, and can have overfitting, which reduce to recognize the original[4]. Not all features help to classify of attacks so choosing the features is an important step . Balance between classification accuracy and feature reduction is research problem, especially if we need to detect multiple types of DDoS attacks"[6] .

To solve this problem, this paper introduces a multiobjective optimization methodology using the Nondominated Sorting Genetic Algorithm II (NSGA-II) for feature selection in multiclass DDoS detection. The suggested framework finds the solutions by maximizing accuracy and minimizing the features at the same time.[2] This improves detection speed, lowers the amount of computing power needed, and improves generalization performance across different types of attacks by combining NSGA-II with a supervised classification model.

II. RELATED WORK

To detect Distributed Denial of Service (DDoS) attacks traditional methods use signature-based and rulebased intrusion detection systems. These methods give a better output to known attack patterns, but they usually miss new and changing attack variations. To overcome the constraints, used machine learning-based Intrusion Detection Systems (IDS), which can identify patterns from network traffic and categorize them. Supervised learning models such as Decision Trees, Support Vector Machines (SVM), Random Forest, and Neural Networks gave better results in identifying DDoS attacks from the datasets[3].

Feature selection is the most important part in Intrusion detection system. High-dimensional datasets make computing more expensive, take longer to train, and raise the danger of overfitting, all of which can lower detection accuracy in real time. The literature has examined numerous feature selection techniques, that includes filterbased methods like Information Gain and Chi-square, wrapper-based approaches utilizing search algorithms, and embedding methods that are incorporated with classifiers. Using these methods data became less complex and help to maintain more accuracy in classification and reduce the feature. 4][5].

In cybersecurity , evolutionary algorithms is an efficient optimization tools for feature selection. Best feature subsets can be selected using Genetic Algorithms (GA), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), and Differential Evolution [6]. Compared to traditional methods, these approaches have flexibility and adaptability. Current research approaches takes feature selection as a single-objective optimization problem, and

focusing on maximizing the accuracy for classification while ignoring the importance of minimization of features.

Multi-objective optimization techniques, is introduced to overcome the limitations. The Non-dominated Sorting Genetic Algorithm II (NSGA-II) was Pareto-optimal solutions . while keeping the diversity solution. This have a balanced trade-off between performance and complexity, in engineering optimization, data mining, and feature selection problems [7]. Because of its ability to maximize both feature reduction and detection accuracy this is ideal for managing high-dimensional network traffic datasets .

Most of the method focus on a single performance metric and binary attack detection. Computationally efficient IDS model is developed by filling the gap which is identified through literature review . NSGAII is used a s a

multiobjective feature selection mechanism to balance the classification accuracy and feature reduction in Multiclass DDoS detection.

The proposed framework aims to improve model generalization across various attack categories, which reduces computational costs, and attack detection is improved by combining evolutionary optimization algorithm with supervised learning [10]. This method is helpful for developing intelligent and scalable intrusion detection systems for modern network infrastructures..

I. PROPOSED WORK

In this work by using the Non-dominated Sorting Genetic Algorithm II (NSGA-II), multiclass DDoS detection combines with feature selection for accurate multiobjective optimization framework is achieved . The primary objective of the proposed work is to achieve a good balance between feature reduction and classification detection accuracy.[11] [12] Most featured subset is selected which determines the most ideal subset of features from the traffic dataset but this is not in the case of traditional IDS techniques.
The work begins with preprocessing the dataset , which includes data cleaning, normalization, and labeling of several attack classes. In this process initial feature set is extracted and encoded as chromosome. A feature subset is represented by each chromosome, and its fitness will be evaluated using two goals: reducing the selected features and maximizing classification accuracy. By using Paretobased ranking and crowding distance methods, NSGA-II iteratively creates better feature subsets with evolutionary operators including selection, crossover, and mutation.

NSGAII is generating Pareto-optimal solutions and it provides various solutions between accuracy and dimensionality. Best feature subset is chosen from different solutions and is used to train a supervised classification model which helps to find several kinds of DDoS attacks. Machine Learning algorithm is used to classifying regular traffic and different types of attacks with the help of selected features This reduce the processing time along with high accuracy detection rate [13][14]. An efficient IDS is possible to create which can be used for currently available network situations in real time suing this method.

The above described framework is evaluated through the performance metrics which includes classification accuracy, precision, recall, and F1-score, as well as count of features selected. The result is compared with before and after optimization to verify how the suggested method works. This proposed works aim is to enhance the detection capability , decrease the model complexity and a scalable solution via combining multi-objective evolutionary optimization with supervised learning.

II. FEATURE SELECTION USING NSGA II

A. Importance of Feature Selection

Most important method is feature selection for a better Intrusion detection system. Since the network traffic dataset have a huge features , particularly for multiclass DDoS detection . Since the dimension of data is high it takes a lot of time for training and is more likely to overfit. if so the detection accuracy will be less. [15]. Feature selection helps to develop more efficient models and , helps to identify the multiple attacks in real time by getting rid of unnecessary and irrelevant features.

B. Chromosome Representation and Initialization

For multi-objective feature selection Non-dominated Sorting Genetic Algorithm II (NSGA-II) is used .A binary chromosome is used to represent each candidate solution, to each gene representing a certain attribute in the dataset[16].Based on the value it is decided to select or not and If the value is "1," it means that the feature is chosen; if it is "0," it means that it is not chosen. Chromosomes are selected random and initial population is generated which have different feature subset. So come the first step in the optimization process.

C. Objective Functions for Optimization

Maximizing classification accuracy and minimizing the number of selected features are the two objectives of NSGA II Algorithms. The difference between regular and traffic attack can be most accurately predicted by choosing the feature subset using NSGA II [17][19].

D. Evolutionary Operations in NSGA-II

Evolutionary operators like selection, crossover, and mutation, NSGA-II is used to enhance the population. Non-dominated sorting is used to order solutions based on Pareto dominance, which makes sure that both objectives are optimized in a balanced way [19]. By using the crowding distance technique diversity among solutions , which prevents premature convergence and helps to search a large area dataset .To make better feature subsets in each new generation is generated using these techniques together.

E. Pareto-Optimal Feature Subset Selection

The Multiple non dominated solutions are the results in pareto front for the optimization process[18]. For each

solution, it has a balance between the classification accuracy and the number of features used. needed feature subset is selected from this set based on the performance and efficiency required for the dataset. the dataset chosen are then used for multiclass classification training.so it reduces the processing time.

I. EXPERIMENTAL SETUP

In this session it explains about the experimental setup that is used for the proposed methodology , NSGA-II–based feature selection framework which is for detecting DDoS attacks on multi classes. The methodology consist of preparing the dataset , preprocessing the process, configuring the parameters for NSGA II and how to choose the evaluation metrics[12].

A.Dataset Description

The dataset used here is CIC-DDoS2019 dataset which is developed by The Canadian Institute for Cybersecurity. It includes both the normal and multiple types of attacks which is generated from a controlled environment. CICFlowMeter is used in the dataset for flow-based network traffic features . Different types of attacks likes TCP-based, UDP-based, HTTP-based, and various volumetric and application-layer attacks in this included in the CIC-DDoS 2019 dataset. So this dataset is the most accurate for evaluating multiclass intrusion detection models[14][15][16].

A.NSGA II Parameter Setting

The NSGA-II algorithm was configured to optimize two conflicting objectives:

1. Maximize classification accuracy
2. Minimize the number of selected features

Parameter	Value
Population Size	50
Number of Generations	100
Cross Over Probability	0.8
Mutation Probability	0.02
Chromosome Encoding	Binary Representation

II. RESULTS AND DISCUSSION

A. In this section it tells about how the proposed NSGAII–based multi-objective feature selection framework works with the CIC-DDoS2019 dataset. the aim is to balance between how accurate is the classification was and how the feature subset is reduced in a multiclass DDoS detection.

B. Feature Reduction Performance

The CIC-DDoS dataset has more than 80 flow-based features. By applying the methodology, NSGA-II optimization, the algorithm created a Pareto front with several non-dominated feature subsets which shows a balance between accuracy and dimensionality.

‘

Metrics	**Value**
Accuracy	96.8%
Precision	95.9%
Recall	96.2%
F1 Score	96.0%
Feature Reduction Ratio	58%
Computational Time	1.8 Sec

Table: Evaluation Metrics

The results show that the proposed framework has a high classification accuracy of 96.8%, which indicates that it can easily identify the difference between benign traffic and various types of DDoS attacks. The precision is 95.9% which represent the model has a low false-positive rate. The recall value of 96.2% indicates that the most attacks are accurately recognized, that is less false negatives. The balanced F1-score is 96.0% it indicates the methodology is working well for all the classes for the classification algorithms.

The results from the experiments confirm that NSGA-II is effectively reconciles two opposing objectives: optimizing accuracy detection and reducing the number of selected features. Single-objective feature selection methods prioritize accuracy than computational efficiency, the proposed multiobjective approach ensures an optimal balance between performance and complexity .By applying the methodology the selected feature subset is enough to discriminate the attacks in CIC-DDOS dataset. By applying the proposed methodology 58% feature subset is reduced with high accuracy and low computational time .So, combining NSGA-II with supervised learning is a reliable and effective way to find multiclass DDoS attacks.

III. CONCLUSION

This research introduced a multi-objective optimization framework for multiclass DDoS detection with the Nondominated Sorting Genetic Algorithm II (NSGA-II), aimed at achieving a balance between classification accuracy and feature reduction[8]. The suggested method solved the problems that come with high-dimensional network traffic data by selecting the most important feature subsets . By combining the NSGA-II with a supervised classification

algorithm the proposed method helps to find both benign and numerous types of DDoS attacks .

In work shows that NSGA-II is a better way to optimize the scalable and accurate multiclass DDoS detection systems. The ability to generate Pareto-optimal feature subsets proved a way to choose solutions based on the needs of the system. The proposed framework can be enhanced in future by integrating deep learning classifiers, real-time traffic analysis, and hybrid optimization strategies to elevate detection accuracy and adaptability in real-world networks[18].

REFERENCES

1. Shubhkirti Sharma, Vijay Kumar, Kamlesh Dutta, Multi-objective optimization algorithms for intrusion detection in IoT networks: A systematic review, Internet of Things and Cyber-Physical Systems Volume 4, 2024, Pages 258-267 ISSN 2667-3452

2. ElDahshan, K. A and , AlHabshy, A. A., and Hameed, B. I. , F.: Meta-Heuristic Optimization Algorithm-Based Hierarchical Intrusion Detection System. Computers 11(12), 170 (2022) https://doi.org/10.1016/j.iotcps.2024.01.003.Reza 3. Prasanna Ranjith Christodoss (University of Technology and Applied Science, Shinas, Oman), Rajesh Natarajan (University of Technology and Applied Sciences, Shinas, Oman), Syed Khaja Mohideen (University of Applied Science and Technology, Salalah, Oman), Justus Selwyn (John Brown University, USA), and B. Anandapriya (Patrician College of Arts and Science, India) , An Efficient Feature Selection in Classifying Diabetes Mellitus, ISBN13: 9781668469712|ISBN13 Softcover: 9798337347219|EISBN13: 97816684697 29 DOI: 10.4018/978-1-6684-6971-2.ch003

4.S. K. Sriramulugari, and V. A. K. Gorantla , and V. Gude, abd K. Gupta and N. Yuvaraj , S: Exploring Mobility and Scalability of Cloud Computing Servers Using Logical Regression Framework," In:2nd International Conference on Disruptive Technologies (ICDT),Greater Noida, India, 2024,pp. 488-493 (2024) https://doi.org/10.1109/ICDT61202.2024.10489115

5.Barradas-Palmeros, J.-A., Mezura-Montes, E., Rivera-López, R., Acosta-Mesa, H.-G., & MárquezGrajales, A. (2024). Computational Cost Reduction in Multi-Objective Feature Selection Using Permutational-Based Differential Evolution. *Mathematical and Computational Applications*, *29*(4), 56. https://doi.org/10.3390/mca29040056

6.Dey, A. K., Gupta, G. P., & Sahu, S. P. (2023). Hybrid meta-heuristic based feature selection mechanism for cyber-attack detection in IoT-enabled networks. Procedia Computer Science, 218, 318–327.

7.Wei Long , Junyu Zhu * , Xuerui Hao , Bin Wu , Chunlin Zhang , Shenghua Liu , Yang Liu , Shengyi Chen , Jian Wu , Xiang Li , Xiao Li * Enhanced Single-Objective Optimization Algorithm with Progressive Exploration Strategy for RF Accelerating Structure Optimization , Posted Date: 25 July 2025 doi: 10.20944/preprints202507.2182.v1

8.C. -Y. Lee and T. -A. Le, "Intelligence Bearing Fault Diagnosis Model Using Multiple Feature Extraction and Binary Particle Swarm Optimization With Extended Memory," in IEEE Access, vol. 8, pp. 198343-198356, 2020, doi: 10.1109/ACCESS.2020.3035081.

9.Yao Zhang, Peng Sha, Meihong Wang, Cheng Zheng, Shengyuan Huang, Xiao Wu, Joan Cordiner,Toward Intelligent and Green Ethylene Manufacturing: An AI-Based Multi-Objective Dynamic Optimization Framework for the Steam Thermal Cracking Process,Engineering,Volume 52,2025,Pages 160-171,ISSN 2095-8099, https://doi.org/10.1016/j.eng.2025.06.045

10. Ms. Preeti Lakhani1*, Dr. Bhavya Alankar1†, Dr. Syed Shahabuddin Ashraf1†, Dr. Suraiya Parveen1†"Harnessing Metaheuristics for Superior Intrusion Detection: Deep Learning on Benchmark Datasets" Journal of Information Systems Engineering and Management 2025, 10(34s) e-ISSN: 2468-4376 https://www.jisem-journal.com/

11. Chandrasekaran, K. P. (2024). Integrating Novel Mechanisms for Threat Detection in Enhanced Data Classification using Ant Colony Optimization with Recurrent Neural Network. Journal of Cybersecurity & Information Management, 14(2).

12. Al Ogaili, R. R. N., Raheem, O. A., Abdkhaleq, M. H. G., Alyasseri, Z. A. A., Alsaidi, S. A. A. A., Alsaeedi, A. H., ... & Manickam, S. (2025). AntDroidNet Cybersecurity Model: A Hybrid Integration of Ant Colony Optimization and Deep Neural Networks for Android Malware Detection. Mesopotamian Journal of CyberSecurity, 5(1), 104-120.

13. Chen, T., & Li, M. (2023). The weights can be harmful: Pareto search versus weighted search in multiobjective search-based software engineering. ACM Transactions on Software Engineering and Methodology, 32(1), 1-40.

14. https://www.kaggle.com/code/eneskosar19/intrusion-detection-system-nsl-kdd

 ISBN: ***978-93-5810-224-6***

An AI-Based Conversational System for Personalized Academic Guidance

Sanjana Maria P S, Shijini Shaji
Department of Cybersecurity and Applied Computing
St.Teresas College ,Ernakulam
Kochi,India
sanjanamaria09@gmail.com,shijinishaji05@gmail.com

Abstract

Academic and career decision-making has become increasingly complex due to the expansion of educational pathways and rapidly evolving industry requirements. In addition to this, traditional academic counseling has limitations due to accessibility and personalization. In this paper, it has been proposed to develop a system based on artificial intelligence technology to provide academic and career counseling to students. In this respect, a conversational AI interface and a machine learning module will be integrated to provide academic and career counseling to students. In this case, the machine learning component will help in the analysis of the students' background, interests, and academic and career aspirations to offer them recommendations. Moreover, with the use of historical data and user experience, the recommendations from this component will be updated depending on the students' profiles. Additionally, based on this proposed model, support for academic and career decision-making, as well as student engagement, has been observed to improve. Moreover, this proposed model ensures the use of conversational AI and machine learning as tools to offer academic and career counseling support.

Keywords— *Artificial Intelligence, Machine Learning, Conversational AI, Academic and Career Counseling, Personalized Recommendations*

I. Introduction

The development of digital technologies and the use of Artificial Intelligence (AI) have significantly affected the education sector. The modern education environment is increasingly shaped by factors such as the diversity of students, the flexibility of the curriculum, and the dynamics of careers. In the modern education environment, which is dynamic and constantly evolving, the need to support students in making decisions on different aspects of education is a key phenomenon that requires constant attention. The current academic support system that enables human teachers to play an important role has provided several limitations that have been apparent, especially in large student populations [1],[3].

Moreover, academic support is identified as one of the most important functions for improving student retention, academic performance, and learning satisfaction [3]. Personalized support helps students to recognize their strengths and weaknesses, align academic choices with career goals, and make use of authentic learning resources at the right time [3]. However, despite its significance, the task of individualized academic advising for all students is a very difficult process due to time and resource constraints. Consequently, students are provided with generalized academic advising, which, in turn, may adversely affect their academic growth [1].

Recent developments in AI technology, especially in the area of Natural Language Processing and Conversational Systems, have opened up promising ways to satisfy the need for intelligent and efficient academic support services. The 'Conversational AI Systems,' commonly referred to

as 'Chatbots,' possess the special capability to understand the conversation of the users and react accordingly.

Conversational AI systems have the potential to operate 24/7 and can provide instant academic assistance without the requirement of human intervention at all times. Apart from the academic performance of individual learners, their interests may vary, and such academic assistance services have the potential to contribute significantly in this aspect. Academic assistance services can be provided to the learners based on the interests of the learners, which is a significant aspect in the present era. In terms of applications, AI-based conversational systems have great potential in the field of education in applications such as automatic tutoring, learning aids, and administrative assistants, among others [8], [10]. Interestingly, studies have suggested that young individuals have a positive mindset in adapting AI chatbots to enhance learning effectiveness and engagement in a higher academic environment [5]. However, AI conversational systems for academic advising as an emerging area of study is an understudied field that needs to be investigated further as an effective academic guidance system not only requires accuracy in recommendation but also covers issues such as fairness, bias, transparency, and sustainability in engagement.

This paper proposes an AI-based conversational model for the purpose of providing personalized academic guidance. This AI-based conversational system aims at providing academic guidance to the students during their academic path. It interacts with the learners through a natural language dialogue, analyzes the relevant academic and personal factors, and provides learners with personalized academic guidance in the domain of academic and career growth. This AI-based system for academic guidance utilizes the capabilities of the AI-driven conversational model for the improvement of the academic guidance process based on recent technological advancements in AI-empowered educational support systems [4], [6]. The rest of the paper is structured in the following sections: Related works in AI-driven educational support systems are presented in Section II, the methodology is elaborated in Section III, the implementation and results are discussed in Section IV, and the conclusion is presented in Section V.

II. Literature Review

In the past decade, however, Artificial Intelligence (AI) technology advancements have had a considerable effect on the education domain, resulting in the development of intelligent systems for assisting, learning, and decision-making for academic purposes. In this context, AI-driven conversational systems have been recognized for their ability to offer support to students. In this section, a discussion on recent and relevant studies regarding AI-driven conversational systems, academic advising systems for higher education, and personalized guidance for students is provided.

To cope with the challenges associated with conventional academic advising, Assayed et al. [1] proposed a systematic review; the paper dealt particularly with conversational AI chatbots used in academic advising. The authors' contributions included a discussion on different architectures, interaction techniques, as well as deployment strategies of different chatbots in learning centers. According to the authors, conversational agents have the potential to significantly alleviate human academic advisors of their responsibilities while providing immediate answers to student inquiries. They also mentioned some drawbacks, including shallow personalization, long-term student modeling, as well as ethical concerns, while emphasizing the need for developing more intelligent conversational academic advising systems.

A systematic literature review by Labadze et al. [2] assesses the applications and role of AI chatbots, taking into consideration its application in learning settings. The systematic literature review included different aspects of AI chatbots, including tutoring, learning support, and student services.

The literature review suggested how AI chatbots can improve student engagement, allowing better accessibility, mainly in an online environment. However, despite the positive attributes of AI chatbots, it is evident from the literature that most AI chatbots function based on predetermined rules or by identifying limited intents.

The role that AI has to play in reshaping academic advising was further investigated by Albinali et al. [3] in their systematic literature review research. The research was centered on intelligent systems geared toward providing academic advising to students. The research revealed various aspects, including that AI technology has great potential in enhancing accuracy in decisions as well as increasing student satisfaction, though most systems lack real-time adaptability as well as integration with student performance information. The research revealed a major research gap concerning personalized conversational systems.

Yusuf et al. [4] have presented a conceptual framework concerning the use of pedagogical AI-based conversational agents in higher education. Additionally, the publication presented a comprehensive survey on the use of existing conversational agents. The publication underlined the importance of aligning the use of conversational agents with pedagogical purposes and learners' understanding in order to enhance the effectiveness of the learning process. The use of conversational agents is common in the learning process to respond to queries; however, only a few agents offer the provision of academic guidance.

Furthermore, student perception and acceptance of AI chatbots have been extensively researched. The study carried out in [5] focused on the mindset of the students in accepting the effectiveness of AI chatbots in facilitating the effectiveness of online learning in higher learning institutions. According to the study, the overall attitude of the student population indicates that they will accept chatbot-based technology as long as they receive accurate, timely, and understandable feedback from these tools.

The nature of the interaction between the students and the educational chatbots was explored by Verleger and Pembridge [6] through a systematic review of the studies on the chatbot interaction process. The study by the authors was useful in understanding the serious role that the nature of the interaction process plays in the students' satisfaction levels with the educational chatbots. As evident from the study conducted by Verleger and Pembridge, the nature of the interaction process to a large extent impacts the trust levels of the student on the educational chatbots.

Ethical factors such as bias and fairness in AI-assisted systems are extremely significant. A study on the use of a bias-aware AI chatbot for engineering advising at the University of Maryland is noted in [7]. The study highlights the significance of the inclusion of fairness-aware mechanisms in conversing systems. The case findings showcased the need to focus on the ethics of AI in coming up with systems to assist with student advising, considering the effect it may have on students' decisions on their future.

The recent advances in the capabilities of generative AI and large language models (LLMs) are further extending the capabilities of conversational systems in learning and education. A study by Elnaffar et al. [8] presents a systematic review on the effectiveness and application of AI chatbots, generative models, and tutoring systems in programming education. According to this study, the potential effectiveness of conversational AI in improving learning and student engagement was revealed, along with the important consideration that the systems must be properly designed to avoid misinformation.

Apart from that, the practical applications of conversational AI in higher learning have been seen in the context of the recent study. For instance, Koutsiaki et al., in their study conducted on the practical application of conversational AI in a higher learning institution setting, introduced a retrieval-based generation chatbot that was aimed to be used in a higher learning setting among Greek-speaking endusers. The practical application of conversational AI was beneficial in the enhanced information access provided to learners. Although the study on the practical application of conversational AI was promising in providing information to Greek-speaking learners in a higher learning setting, there were certain shortcomings observed from the study in providing learners with a whole range of academia information. Recent advancements in generative AI models and large language models have greatly enhanced the practical applications of conversational AI in the education sector. Within the recent advancements in the sector, a study

on the practical application of AI chatbots, generative models, and tutoring tools in the education sector for programming was conducted by Elnaffar et al. [8]. The study concluded that the incorporation of conversational AI in a higher learning institution setting can be beneficial in enhancing learning outcomes.

Recent research has also highlighted several other applications of conversational AI in the context of higher education. Koutsiaki et al. [9] suggested a case study on the application of an RAG-based conversational AI system in Greek language higher education. The proposed system improved the availability of information and thus increased student engagement through the application of conversational AI to institutional knowledge bases. The study did, however, suggest that the proposed approach, while extremely effective in the context of informational queries, has not yet reached the level of holistic, personalized academic counseling for very diverse student profiles.

Abedi et al. [10] explored the possibilities of large language model-based conversational AI systems in graduate engineering programs. The study highlighted the ability of the LLM-based conversational AI system to address complex academic queries, provide personalized feedback, and facilitate advanced learning tasks. The authors suggested that while the proposed systems have immense possibilities, the application of these systems to structured academic counseling would require a deliberate convergence of academic rules, student information, and ethical considerations.

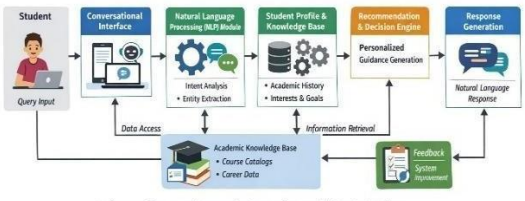
From the literature reviewed, it is evident that the conversational AI technology has been proven to possess immense potential for the improvement of academic support services. However, based on the studies conducted on this topic, there exist certain limitations in terms of personalization, adaptability, ethical transparency, and academic guidance. These limitations led to the development of an AI-based conversational technology for the provision of personalized academic guidance, as discussed in this study, which aims to integrate conversational intelligence, student modeling, and ethical AI for efficient academic advisory support.

III. Methodology

The prime objective of the current research is to create a methodology for developing AI-based conversational systems for providing academic guidance services to students in higher education. In this context, some researchers and studies pointed out that academic advising services can be efficiently supported by conversational AI systems with the potential for scalability and accessibility along with reduced human dependency [1], [3]. The proposed methodology for the development targets academic advisory services by utilizing conversational intelligence along with profiling and recommendation capabilities.

A. Proposed Methodology

Fig. 1 illustrates the schematic representation of the proposed AI-based conversational system for personalized academic guidance.



Architecture of the AI-Based Conversational System for Personalized Academic Guidance

Fig. 1. Architecture of the AI-Based Conversational System for Personalized Academic Guidance

The proposed methodology is primarily centered on the provision of academic guidance that is studentcentric through natural language interactions. The students are encouraged to submit queries related to the selection of different courses, academic development, skill development, or even career development through natural language interaction systems. Conversational interaction systems have been widely recognized for the provision of continuous interactive educational support.

The workflow involved in the system would include query understanding, student profiling, recommendation generation, along with the delivery of the response. An analogous application of the architecture has been employed for the design of chatbots for academic guidance, resulting in added responsiveness and the ability for more personalization [1], [4]

B. Student Data Gathering & Pre-processing

The need for personalized academic counseling implies the necessity for deep knowledge regarding individual learner characteristics. The proposed system relies on a student profile repository that maintains academic and contextual information such as academic history, course performance, interests, learning preferences, and career aspirations.Previous works emphasize that appropriate student modeling is one of the significant assumptions on which beneficial academic advice can be offered to students [3], [4].

Student data can be gathered from institutional repositories, questionnaires during the induction process, and dialogue interactions. Preprocessing methods address the issue of incompleteness and inconsistencies in academic data. Adaptation and long-term academic support rely on the continuous update of the student model, which is a drawback in current state-of-the-art systems [1], [3]

C. Natural Language Processing, Intent Classification

NLP is the key to the conversational intelligence that powers this system. The arriving queries will go through some basic preprocessing operations such as tokenization, stop-word removal, and lemmatization. The intent classification methods will then be able to classify the type of student query, whether it is related to academic counseling, course suggestions, or career guidance. The use of NLPpowered conversational interfaces has been shown to enhance the quality of interactions and student engagement in higher education settings [2], [6].The entity extraction mechanisms are tasked with extracting important academic terms, course names, academic levels, and domains of interest. Advanced conversational models, including large language models, can be integrated to support multiturn dialogue and contextual understanding, as suggested in recent studies on pedagogical AI systems [4], [8], [10].

D. Customized Recommendation and Decision Engine

Once the student's intention and relevant entities are identified, the system provides customized academic guidance by leveraging a recommendation and decision engine. This engine combines rulebased academic constraints with data-driven decision mechanisms to provide accurate and institutionspecific recommendations. Previous works in literature emphasize that integration of academic rules with student data will help advise with greater accuracy [1], [3].

The decision engine evaluates multiple academic pathways based on their student performance, interests, and long-term goals. Ensuring ethical considerations like fairness and mitigation from bias will keep recommendations fair and will address most of the issues raised in bias-aware academic advising systems [7].

E. Knowledge Base Integration

The system will include an academic knowledge base that will consist of curriculum structures, course catalogs, institutional policies, skill sets, and career paths. Knowledge-based

support is fundamental in ensuring proper academic support that is context-sensitive and reliable [2], [9].

To make response accuracy better, retrieval-augmented generation methods can be used. These methods will enable the chat system to obtain verified academic information before responding. This has already been proven to improve the transparency of conversational AI systems and reduce misinformation [9], [10].

F. Response Generation and Feedback Mechanism

The response generation module takes the recommendations and provides a response to these recommendations, which is clearer and easier for students to understand, as well as concise. The clarity and usability of conversational responses are highly important and crucial for students' acceptance of an artificial intelligence-based academic guidance system [5], [6].

A feedback system is employed to examine system efficacity and user satisfaction. Feedback is applied to optimize recommendation strategies and enhance dialogual quality. Continuous learning and development are identified as critical features for the long-term accomplishment of AI-aided assisting systems [4], [8].

G. Workflow of the Proposed System

The entire workflow of the proposed system can be defined in the following manner:

- Student submits a query through the conversational interface.
- NLP module process the input and identifies the intent & entity
- Student profile and academic knowledge base are accessed
- Personalized academic guidance is produced by recommendation engine
- Response is given in natural language
- Feedback collection for system improvement

The proposed methodology ensures scalability, personalization, and ethical academic guidance by incorporating conversational AI with student-centric decision-making methods, which appropriately addresses the limitations presented in the current student academic advising systems [1], [3], [7].

IV. Implementation and Discussion

In this study, multiple experimental evaluations are carried out to assess the efficacy and effectiveness of the proposed AI-based smart education chatbot system. The proposed system is evaluated in terms of the accuracy of the responses, the ability to personalize the responses, and the overall quality of the interaction with the users.

A. System Implementation

The proposed smart education bot system is implemented in the form of a prototype system that uses chatbots to help a student with academic-related guidance. The components of the system include a user interface module, natural language processing module, recommendation system, and knowledge base.

The student queries handled by the chatbot involve natural language processing tasks such as tokenization, intention identification, and entity extraction. After the intention has been determined, the relevant academic details are obtained from the knowledge base, while the personalized recommendations are generated electronically by the formation of rules based on academics and students. It can have multiple-turn conversations for clarity.

B. Query Processing and Response Generation Results

The chatbot system was tested using a set of sample academic queries related to course selection, academic plans, skill acquisition, and career guidance. The chatbot was successful in understanding most user queries and responding accordingly.

The results revealed that the system was effective in providing precise academic advice on student common queries. In addition, the query relating to course prerequisites and planning received fast and precise replies, as opposed to the general academic advisory system. The time taken by the system to generate the response was minimum.

C. Performance Evaluation

The performance evaluation of the proposed system is carried out through various qualitative parameters. It has been noted that the proposed chatbot showed performance for various categories of queries.

Future work can take advantage of the personal responses based on the students' interests and educational background. The need for repetitive questions was clearly diminished, and the structured guidance was highlighted as useful for the academic advising role.

D. Discussion

Based on the experimental result, the proposed smart education bot has shown its potential to provide personalized and timely educational guidance. In addition, integration with conversational AI enables more student engagement and facilitates timely resolution of academic uncertainties.

Nevertheless, the system's performance is greatly influenced by the overall quality of the academic knowledge base as well as the accuracy of the information provided by the student. Limited real-world deployment and the dependency on pre-defined intents are major drawbacks. A possible addition in the future is the development of intelligent learning capabilities along with voice interaction.

Overall, the proposed system has significant potential for being an intelligent academic assistance system, which would lead to better support for students and decision-making in education.

V. Conclusion

The current research introduced an AI-based conversational model with the capability of delivering accurate personalized academic support to students at the higher educational institution level by integrating natural language processing and the concept of a recommendation system. Such a conversational system can help the student in many ways, such as course selection, planning, skill building, and more, thereby reducing dependency on academic support systems.

The experimental testing of the prototype has shown that the system is capable of understanding the questions being asked by the students correctly and answering them accordingly. From the obtained results, it is clear that the proposed approach is helpful in achieving higher accessibility, scalability, and personalization in academic counseling. Additionally, the system also emphasizes transparency so that the suggestions made can be easily comprehended.

VI. FUTURE SCOPE

Although the proposed system holds promises of new performance, at this stage, the system is reliant on predetermined intents and academic data structure. The future development of the system may focus on the inclusion of more intelligent models of language to enhance the level of conversation and the learning ability of the system.

Other improvements may also include the inclusion of voice interaction, real-time integration with the academic management system of the institution, and the use of recommendation systems based on the available data. The issues of fairness, transparency, and data privacy will also become important for the proper use of the system.

REFERENCES

[1] S. K. Assayed, M. Alkhatib and K. Shaalan, "A Systematic Review of Conversational AI Chatbots in Academic Advising," in Lecture Notes in Civil Engineering, vol. 473, Springer, Cham, 2024.

[2] L. Labadze, M. Grigolia and L. Machaidze, "Role of AI chatbots in education: systematic literature review," International Journal of Educational Technology in Higher Education, vol. 20, art. no. 56, 2023.

[3] K. Albinali, N. A. Iahad, A. F. Yusof and A. Abdulla, "A Systematic Literature Review of the Artificial Intelligence Role in Transformative Academic Advising," International Journal of Academic Research in Business and Social Sciences, vol. 14, no. 12, 2024.

[4] H. Yusuf, A. Money and D. Daylamani-Zad, "Pedagogical AI conversational agents in higher education: a conceptual framework and survey of the state of the art," Educational Technology Research and Development, vol. 73, pp. 815–874, 2025.

[5] "Students' mindset to adopt AI chatbots for effectiveness of online learning in higher education," Future Business Journal, vol. 11, art. no. 30, 2025.

[6] L. F. Verleger and T. Pembridge, "Interacting with educational chatbots: a systematic review," Education and Information Technologies, 2022.

[7] "Bias-Aware AI Chatbot for Engineering Advising at the University of Maryland A. James Clark School of Engineering," arXiv, 2025.

[8] S. Elnaffar, F. Rashidi and A. Z. Abualkishik, "Teaching with AI: A systematic review of chatbots, generative tools, and tutoring systems in programming education," arXiv, 2025.

[9] M. E. Koutsiaki et al., "From Textbook to Talkbot: A case study of a Greek-Language RAG-based chatbot in higher education," arXiv, 2025.

[10] M. Abedi, I. Alshybani, M. R. B. Shahadat and M. S. Murillo, "Beyond traditional teaching: the potential of large language models and chatbots in graduate engineering education," arXiv, 2023.

[11] A. Følstad and P. B. Brandtzaeg, "Chatbots and the new world of HCI," Interactions, vol. 24, no. 4, pp. 38–42, 2017.

[12] A. Winkler and M. Söllner, "Unleashing the potential of chatbots in education: A state-of-the-art analysis," in Proc. Academy of Management Annual Meeting, 2018.

[13] M. Okonkwo and A. Ade-Ibijola, "Chatbots applications in education: A systematic review," Computers and Education: Artificial Intelligence, vol. 2, 2021.

[14] E. Adamopoulou and L. Moussiades, "An overview of chatbot technology," in Proc. IFIP International Conference on Artificial Intelligence Applications and Innovations, 2020, pp. 373–383.

[15] P. Smutny and P. Schreiberova, "Chatbots for learning: A review of educational chatbots for the Facebook Messenger," Computers & Education, vol. 151, 2020.

[16] C. Troussas, M. Virvou, K. J. Espinosa, K. Llaguno and J. Caro, "Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning," in Proc. IEEE IISA, 2013.

[17] S. K. D'Mello and A. Graesser, "AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers," ACM Transactions on Interactive Intelligent Systems, vol. 2, no. 4, 2012.

[18] Z. Chen, H. Wu, and L. Zhang, "Artificial intelligence in education: A review," IEEE Access, vol. 8, pp. 75264– 75278, 2020.