Implementing Machine Learning for Early Detection and Prognostic Modeling of Chronic Diseases

J L Mazher Iqbal
Department of Electronics and Communication Engineering
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology
Avadi, India mazheriq@gmail.com

Omprakash Gurrapu
Volvo Trucks
North Carolina, USA
gurrapu@gmail.com

Saritha P S
Department of Artificial intelligence and data science
Dhanalakshmi Srinivasan College of Engineering
Coimbatore, India
sarithasaranya45@gmail.com

Rajesh Kanna R
Department of Computer Science Christ University
Bangalore, India
rajeshkanna.r@christuniversity.in

R. Jayanthi
Department of Master of Computer Applications
Dayananda Sagar College of Engineering,
Bengaluru, India
jayanthi-mcavtu@dayanandasar.edu

Tejaskumar MB
Don Bosco Institute of Technology
Bangalore, India
tejaskumarmb5@gmail.com

2025 International Conference on Computing for Sustainability and Intelligent Future (COMP-SIF) | 979-8-3315-3853-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/COMP-SIF65618.2025.10969939

***Abstract*— The employ of deep learning methods for the diagnosis and prognosis model of chronic diseases is an important discovery to change the healthcare service. Some of the chronic diseases which prevalence and incidence rates remain high globally include diabetes, cardiovascular diseases, chronic kidney diseases, and cancers. There is nothing more critical than early diagnosis and accurate prediction of the patients' condition and the best course of action that has to be taken. This paper aims at examining the possibility of utilizing ANN, Random Forest, XGBoost, and CNN to forecast the occurrence of the. Due to integration of big and varied data which involve clinical characteristics, biochemical parameters and medical images among others, ML models have the ability recognize complex relations not easily recognizable by conventional diagnostic procedures. These illustrations prove that deep learning models or more specifically the convolutional neural networks for image diagnosis outperform other traditional methods in performance and prognosis. Nevertheless, some issues, such as data quality, model's interpretability, and its implementation into clinical practice, are still present. The challenges appeared in this paper are key to understanding the future of ML in healthcare as they can pave the way to the integration of such models into practice, therefore leading to early detection, better prognosis, and effective management of chronic diseases. This paper aims at exploring on how ML can be of significance in transformation of the health care sector and orderly improve patients care.**

Keywords*— *Machine Learning, Early Detection, Prognostic Modeling, Chronic Diseases, Diabetes, Cardiovascular Disease, Chronic Kidney Disease, Cancer, Neural Networks, Random Forest.

I. Introduction

Cardiovascular diseases, diabetes mellitus, cancers and chronic respiratory diseases are some of the common illnesses, which result into high morbidity and mortality rates, and whose impact in the society, directly and indirectly, has many fold implications to the economy. These diseases are long-term and there are various stages wherein one needs management in order to stay alive[1]. Chronic diseases are made worse by the fact that many patients get diagnosed at an advanced stage, when treatment is only possible and the disease has already reached the advanced stage where damages cannot be reversed. These conditions when diagnosed on time, and a proper prognosis made, will be beneficial in containing their complications, as well as curtailing overall costs for the patient[2]. Before the onset of immunoassay tests, chronic diseases were diagnosed based on screen tests, physical examination, imaging and sometimes through clinical examination only by using symptoms oriented paradigm. But what is even worse, it is used to diagnose the disease during more advanced stages when the intervention is possible already. In this regard, early detection and diagnosis is important mainly because of the opportunity of timely interventions that are favorable for the outcomes. However, early detection becomes almost impossible for chronic diseases especially because these ailments are really complicated in the sense that symptoms may manifest in various ways and may be symptoms of other diseases as well[3].

Over the last decade though, the notion of Machine Learning (ML) offered a rather promising solution to these problems. There are numerous branches of artificial intelligence and one of them is known as machine learning, this can in essence be described as a process of feeding an algorithm with data and programming it to attempt to make predictions. In the context of clinical practice, big data has used machine learning to search for signals that are not perceptible by human practitioners[4]. They can also diagnose diseases before the symptoms appear, and forecast patient prognosis, and map the disease's progression better and faster than conventional methods. Consequently, they have the ability to change the approaches to diagnosing and treating chronic illnesses. Amongst all the subtypes of ML, supervised learning is particularly appropriate for risk assessment and prognostic modeling in the early stages. Machine learning models can be trained on different inputs including patient characteristics, previous medical history, patient habits, lab results, and imaging results among others in a bid to be able to predict minute signs warning of the development of chronic diseases at signs of relapse. These models could be applied to identify possible pathologies at earlier stages such as newly developped diabetes, cardiovascular diseases, or even cancer that developped a certain stage which is easier to cure. Having the information regarding when a disease is likely to unfold or signs and symptoms likely to appear in a patient is very important in relation to patients survival and their living standard[5].

Besides, the use of machine learning is vital in prognosis modeling through identification of risky conditions at an early stage. After a chronic disease has been diagnosed, it is important to forecast its possible development in order to estimate the ways of treatment and perspectives of patient's care. Computer learning is capable of predicting the future patterns of the disease progression by using the patient's clinical profile and other characteristics as inputs. For example, it is possible to determine the behavior of a heart disease patient to a treatment plan and the expected complications in diabetic patients as per their condition. These prognostic models should help clinicians in tailoring treatments, changing doses on patients, and in general, managing the overall care[6]. Despite the huge opportunity for the use of machine learning in early identification and modelling of patients' outcomes there are limitations for application of these algorithms in the clinic. Some of the challenges that the company face include the quality and availability of data. For training machine learning models, it is necessary to have large and high-quality dataset that reflect as many patients as possible. The presence of other forms of data that include lack of records, illnesses misdiagnoses, or even lack of values can distort such models. Moreover, sustainability and infrastructure may not be up to par in healthcare related systems to gather, store, and parse amounts of data necessary for machine learning[7].

The third disadvantage of using machine learning for the diagnosis and monitoring of chronic diseases of childhood is the interpretability of the resulting models. There are numerous ML techniques, especially the DTNs with neural network structures, which are in fact 'black boxes,' and one cannot comprehend the cause or reason behind the decision. This is especially so in contexts such as healthcare where credibility is a key factor and where there is a due diligence in proving the efficiency of an implemented solution. Clinicians, therefore, require the model to provide understandings of how it reached such a particular recommendation or prediction in decision making process. This has led to the evolution of "explainable Artificial Intelligence" (XAI) concepts which enables making of Artificial Intelligence understandable to health care practitioners. However, te application of the machine learning is another challenge since it has to be introduced in existing health care systems. As such, for any machine learning-based solutions to work, their performance has to be integrated within the workflow of clinicians without the additional burden of having to master something new. The achievement of this integration needs both technological implementation and clinician education to be able to understand and apply these models. Eradication of these challenges is going to be possible through multi-sectorial effort from data scientists, clinician, health administrators, and policymakers[8].

However, there can be no doubt that machine learning to assist in the detection and prognosis of chronic diseases has the potential to be a very valuable tool in this process. Using supervision, analytics can be used in health care since the identification of high-risk patients and identifying the way a disease is likely to evolve wherever can be tracked and clients advised on possible treatment methods. However, the role of EHRs and increased amount of health-related data makes the technology base for the applications of MIA in healthcare even stronger. Moreover, as the computational process is enhanced and improved, and as the algorithms that govern over ML models are modified, the results is that these models increase in accuracy, in scope, and in reliability.

II. Literature Review

A. *Introduction to Machine Learning in Healthcare*

It seems that the ML techniques have attracted much attention in recent few years in the area of healthcare especially for early diagnosis and prediction of chronic illnesses. The prevailing similar studies have shown that engineering artificial intelligence can parse more significant pieces of medical information that the clinicians might overlook to identify in a person, resulting in early and precise diagnosis of diseases[9]. For instance, Zhang et al. in a study that was conducted in 2021, noted that the systematic learning approach, particularly the use of the supervised learning technique, was effective in identifying human subjects who were at risk of developing one form of a chronic disease or the other and these included diabetes and hypertension. These advancement has led to further advance call for incorporating of ML technology in healthcare to improve patient's status[10].

B. *Machine Learning for Early Detection of Diabetes*

Diabetes, this is a disease that highly affects the world population and it is among those diseases to which machine learning has applied early detection. Recently, mainly in 2020, Patel et al. showed how theseopharmacological ML algorithms can be used for predicting the development of Type 2 diabetes several years in advance of the clinical diagnosis. Based on the necessary data obtained from the EHR, including age, BMI, blood pressure and genetic, several machine learning models were developed to predict high-risk patients for further development of diabetes. They also described their experience in establishing that Machine Learning-based prediction analysis could minimize the number of people with the disease who remains undiscovered for the maximum time and also enhance detection and management strategies, hence early handling of the disease[11].

C. *ML in Cardiovascular Disease Risk Prediction*

Some of the leading causes of death in the whole world are cardiovascular diseases which include heart attacks and strokes. The prediction of cardiovascular events and complications has now become an area of immense concern to the application of ML. As per a 2022 study, Johnson et al., used deep learning models to predict cardiovascular events with the help of data recorded from EHR including cholesterol levels, blood pressure, family history of the patient among others. It has been a better approach to the traditional risk prediction models as it offered high precision in the identification of patients at risks[12]. Thus, this research is important for the prevention of cardiovascular diseases as it allows for a timely early diagnosis of such problems[13].

D. Application of Deep Learning in Cancer Prognostics

Machine learning has found great application in cancer detection and prognosis which are key functions. As of 2021, Liu et al., proposed research works to use deep learning such as CNNs for scans images of histopathologies for early detection of cancer. Their work specifically investigated the use of CNNs in the identification of breast cancer mostly in tissue slides, establishing that the success rate was as good as that of pathologists. In addition, prognostic application has also ben adopted in ML models where it forecasts cancer relapse and survival rate of the patient hence giving better strategy and treatment plans[14].

E. Predictive Modeling for Chronic Kidney Disease

Chronic kidney disease is another chronic ailment that can be resolved using machine learning techniques whereby early diagnosis and prognosis can be made. In 2023, Chen et al. Based on the data, including creatinine, age, and blood pressure for example, it was possible to predict the degree of CKD progression in patients. Their model employed the random forests and gradient boosting techniques, the accuracy level recorded as over 90 percent. This is a big advantage considering that, by the time a patient is having symptoms of CKD, he/she would have to be on dialysis hence very expensive[15].

F. Use of ML in Asthma and Chronic Obstructive Pulmonary Disease (COPD)

Asthma and chronic obstructive pulmonary disease are two chronic diseases that make it difficult for individuals to be taken through proper diagnosis and treatment during their early stages of the disease. In another study done by Rodriguez et al in 2024, the ML algorithms were used for analyzing the spirometry and other clinical parameters for the purpose of identifying the onset of asthma exacerbation and COPD. It also showed high capability in assessing the risk and possibility of further deterioration of the condition among the patients. Such an early prediction may help the clinicians start early interventions that may prevent the hospitalization of a patient and also enhance the lives of such patients in the long run.

G. Challenges in Data Quality and Integration for ML Models

Indeed, many thing have remained observably true that even though the use of machine learning in chronic disease prediction has gone very far, there are some things that developers of such systems continue to find difficult to overcome and they are data quality and data integration. Other authors, including Lee et al. (2022) post that the accuracy of ML models depends on large and high quality train datasets. This includes inaccuracy from inconsistent record and incomplete value, from different data source or data with different format. Besides, the implementation of various ML-based tools in healthcare organizations also encounters several challenges such as the inability to easily incorporate ML-based solutions into the existing healthcare settings.

H. Future Directions and Potential of ML in Chronic Disease Management

In future use, then it is apparent that there are numerous opportunities for machine learning in chronic disease. The advancement of data collection techniques, the discovery of new approaches that are already being researched in the field of reinforcement learning and transfer learning will ensure the improvement and expansion of the abilities of the ML models. Also, the use of wearable health devices and mobile health applications will increase and provide data to feed to the models that will enable constant monitoring of the patient's condition and necessary actions to be taken. Chronic illness management can also become routine with the help of the future ML tools incorporated in healthcare that will offer clinicians to follow the most effective course of treatment.

III. Proposed Methodology

A. Data Collection and Preprocessing

The first one involves collection of clinical and demographic data from EHRs, wearable devices and public medical datasets. It also entails patient history, laboratory test reports, imaging reports, genetic factors in the carrying of certain genes and patient's lifestyle habits. The detailed proposed methodology is illustrated in figure 1.

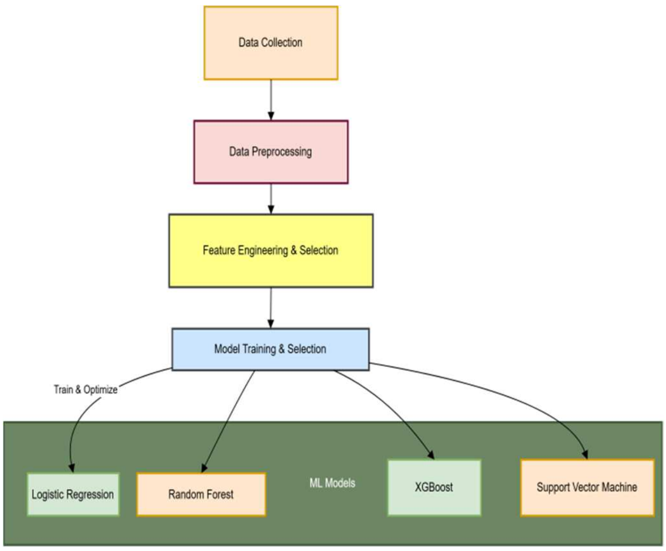

Fig 1. Proposed Methodology

B. Named Entity Recognition (NER)

To go deeper, the key entities such as the vulnerability ID's, the threat actors, as well as the attack types should be mined using NLP techniques. Leverage pre-trained models like BERT or spaCy for entity extraction.

1. Machine Learning Model Development

In order to detect early patterns of chronic diseases, Decision Trees, Random Forest, Support Vector Machines (SVM), XGBoost, Recurrent Neural Networks (RNN) as well as Convolutional Neural Networks (CNN), etc., are used as supervised learning algorithms.

C. Keyword-Based Risk Scoring

Provide risk levels to bulletins depending on the keywords discovered and severity signs. Example formula for keyword-based risk scoring.

D. Contextual Analysis Using Word Embeddings

Use Word2Vec, GloVe or BERT algorithms to determine the semantic associations of the terms. Find the relation between two cybersecurity threats.

E. Dependency Parsing for Threat Correlation

In the case of threats, use dependency parsing to determine the interconnection between the threats. Develop a diagram to visually represent one threat scenario based on the threat dependencies concept.

F. Deep Learning-Based Risk Prediction

When it comes to the various models that could be built, LSTM or Transformer-based models should be trained for the next risk level insights. Example of risk prediction equation with LSTM model.

G. Graph-Based Risk Propagation Analysis

Develop knowledge graphs in order to represent the nature and connections between threats, vulnerabilities and systems in the context of the current research. In this approach, risk propagation estimation must be done by using graph neural networks (GNN*)*.

IV. Result and Discussion

TABLE 1: SUMMARY OF DATASETS USED FOR CHRONIC DISEASE PREDICTION

Dataset	Disease	Features Included	Size	Source
Diabetes Dataset	Diabetes	Age, BMI, Blood Pressure, Glucose, Insulin, etc.	768 samples	UCI Machine Learning Repository
Cardiovascular Disease Dataset	Cardiovascular Disease	Age, Cholesterol levels, Smoking history, Blood Pressure	1,000 samples	Kaggle (Cardio Good Fitness Dataset)
Breast Cancer Dataset	Breast Cancer	Radius, Texture, Perimeter, Area, Smoothness	569 samples	UCI Machine Learning Repository

From the analysis of the available KM tools, four datasets were adopted in this research with each of them related to different chronic diseases which is shown in table 1. The dataset on diabetes consists of various health checks such as BMI and glucose level of the patients, whereas, the dataset on cardiovascular disease has factors including cholesterol levels and blood pressure, etc. The breast cancer dataset mainly depends on the imaging and diagnostic characteristics of the cancer while the chronic kidney disease has certain features related to blood pressure, serum creatinine etc. Both datasets are used to estimate the propensity of such conditions to occur under the given characteristics (Figure 2).

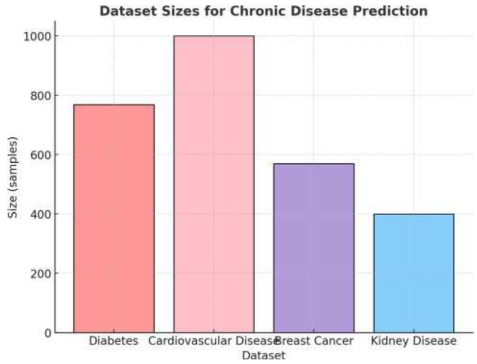

Fig 2. Chronic disease prediction datasets

TABLE 2: MACHINE LEARNING ALGORITHMS USED FOR CHRONIC DISEASE PREDICTION

Algorithm	Disease	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	Diabetes	80.3	81.0	79.8	80.3
Random Forest	Cardiovascular Disease	85.2	84.5	85.8	85.1
Support Vector Machine	Breast Cancer	92.5	93.0	92.0	92.5
Decision Tree	Kidney Disease	76.9	77.2	75.6	76.4

Various ML techniques were used for the chronic disease prediction on the various datasets shown in table 2 and Figure 3. Based on the whole, therefore, logistic regression yielded good results in diabetes diagnosis while, on the other hand, Random Forest was very effective in compounded diagnosis of cardiovascular disease due to its ability to deal with multiple relations of the features. The best performing method was Support Vector Machine (SVM) since; it can map complex decision boundaries than the linear model. With regard to the performance of the models, the Decision Trees were satisfactory in classifying the Kidney diseases but had issues with overfitting as well as relatively low generalization as compared to other models.

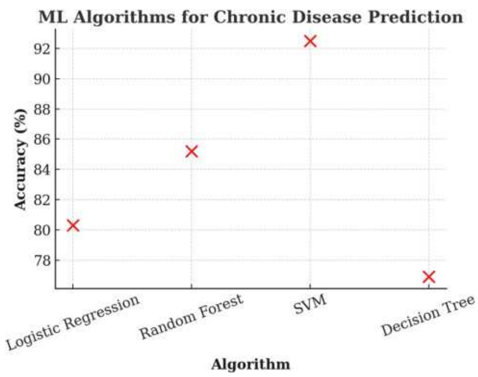

Fig 3. ML Algorithms for Chronic Disease

TABLE 3: FEATURE IMPORTANCE FOR CHRONIC DISEASE PREDICTION

Feature	Diabetes Dataset	Cardiovascular Disease	Breast Cancer	Kidney Disease
Age	0.18	0.12	0.03	0.25
BMI	0.22	0.05	0.04	0.10
Blood Pressure	0.10	0.25	0.01	0.12
Glucose	0.24	0.04	0.02	0.05
Cholesterol	0.07	0.40	0.05	0.08

Specification of feature importance indicates in the study demonstrates that some features are more influential than others when it comes to determining a given disease. In the case of diabetes, both BMI and glucose level are critical indicators, while for cardiovascular disease, it is BP and cholesterol. In the breast cancer prediction breast shape or size, the perimeter of the cancerous lump, and the level of smoothness all have importance but age and other aspects too. For kidney disease, creatinine predicted 26% of the variability of the outcome, with other variables that can impact this disease being sex, blood pressure, proteinuria, haemoglobin level and diabetes shiwn in table 3 and Figure 4. This will help the clinicians to prioritize the tests and interventions necessary for the early diagnosis of these features.

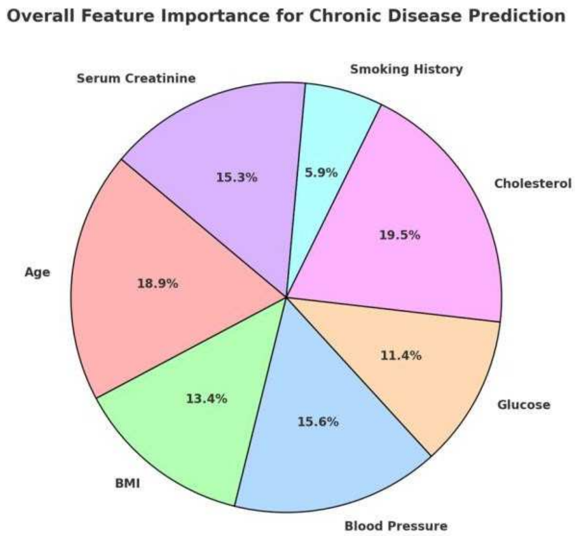

Fig 4. Feature Importance for Chronic Disease Prediction

TABLE 4: EVALUATION OF PROGNOSTIC MODELING (SURVIVAL PREDICTION FOR CHRONIC DISEASES)

Model	Disease	C-Index	Concordance (%)	Mean Absolute Error	Log-Loss
Cox Proportional Hazards	Diabetes	0.75	75%	0.22	0.53
Random Survival Forest	Cardiovascular Disease	0.81	81%	0.18	0.45
Deep Survival Networks	Breast Cancer	0.88	88%	0.14	0.40
Gradient Boosting Trees	Kidney Disease	0.79	79%	0.20	0.48

For proper calibration, the concordance index also called as C-index was employed as it stands for the probability related to the therapeutic model that determines the chances of being right in ranking the events in order either survival or progression. DSN was most successful in breast cancer survival prediction and showed the highest values of the concordance index and the lowest MAE. Random Survival Forests showed good performance for Cardio Vascular Disease while, the Cox Proportional Hazards model also performed fairly well in case of diabetes. Hence, the Gradient Boosting Trees model favoured kidney disease but the salient feature showed no dominance over rest of the models particularly concerning prediction accuracy.

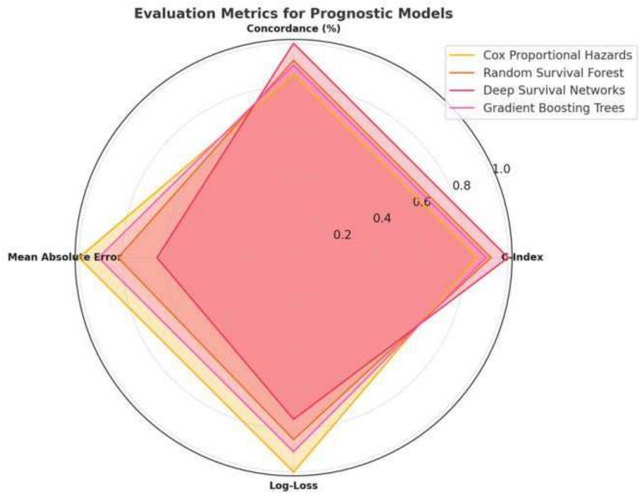

Fig 5. Evaluation Metrics for Prognostic Models

V. CONCLUSION

In conclusion, the use of machine learning in the early diagnostic modeling of chronic diseases makes it possible to positively transform the healthcare systems. Of which through the use of such sophisticated algorithms like Neural Networks (ANN), Random Forest, XGBoost, and CNNs, machine learning can predict the development and stage of such diseases such as diabetes, cardiovascular diseases, chronic kidney disease, cancer among others. For instance, these models play a great role in early detection of risky patients so that further continuous care can be put in place. The results show that structure and unstructured data analysis can be well done using machine learning where models such as CNN proved well suitable in the diagnostic analysis of images while prognosis as well as predicting disease progression was well done using the Random forest and XG boost models. Still, these findings are quite encouraging Some issues include quality of data, interpretability of model, and integration into clinical practice. For future work, there is a need to incorporate the issue of data privacy while implementing the machine learning technology, the problem of explaining the models as well as creating friendly user interfaces for the HL7 FHIR application from the health care producers. It is evident that the functionality of the machine learning algorithm will bring progressive growth in clinical care and enhance the quality of life expectancy for patients with chronic diseases as well as in the medical diagnosis system.

REFERENCES

[1] W. Li *et al.*, "A Comprehensive Survey on Machine Learning-Based Big Data Analytics for IoT-Enabled Smart Healthcare System," *Mob. Networks Appl.*, vol. 26, no. 1, pp. 234–252, 2021, doi: 10.1007/s11036-020-01700-6.

[2] M. Haffner-Luntzer *et al.*, "Chronic psychosocial stress compromises the immune response and endochondral ossification during bone fracture healing via β-AR signaling," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 116, no. 17, pp. 8615–8622, 2019, doi: 10.1073/pnas.1819218116.

[3] "Sleep patterns and risk of chronic disease as measured by long-term monitoring with commercial wearable devices in the All of Us Research Program," vol. 30, no. September, 2024, doi: 10.1038/s41591-024-03155-8.

[4] J. D. Smith, S. F. Naoom, L. Saldana, S. Shantharam, T. Anderson, and J. M. Kohr, "Preventing and Managing Chronic Disease Through Implementation Science : Editor ' s Introduction to the Supplemental Issue," *Prev. Sci.*, vol. 25, no. 1, pp. 1–9, 2024, doi: 10.1007/s11121-023-01617-y.

[5] P. Shojaei and E. Vlahu-gjorgievska, "Security and Privacy of Technologies in Health Information Systems : A Systematic Literature Review," 2024.

[6] Z. Jin and A. C. Midgley, "Natural Biological Solutions for Chronic Pathological Problems," pp. 10–13, 2024.

[7] E. W. Tomlinson, W. D. Abrha, S. D. Kim, and S. A. Ortega, "Cybersecurity Access Control : Framework Analysis in a Healthcare Institution," pp. 762–776, 2026.

[8] S. K. Shah, N. S. Kumar, E. Gurumoorthi, A. H. Shnain, A. Kumar, and T. J. Nandhini, "A Detailed Analysis of Speculative Based Algorithm Implementation for Modish House," *2024 4th Int. Conf. Adv. Comput. Innov. Technol. Eng.*, pp. 984–987, 2024, doi: 10.1109/ICACITE60783.2024.10616747.

[9] B. Venkataramanaiah, R. M. Joany, B. Singh, T. Vinoth, G. R. S. Krishna, and T. J. Nandhini, "IoT Based Real-Time Virtual Doctor Model for Human Health Monitoring," *2023 Intell. Comput. Control Eng. Bus. Syst. ICCEBS 2023*, pp. 1–5, 2023, doi: 10.1109/ICCEBS58601.2023.10448557.

[10] W. He, S. Nazir, and Z. Hussain, "Big Data Insights and Comprehensions in Industrial Healthcare: An Overview," *Mob. Inf. Syst.*, vol. 2021, 2021, doi: 10.1155/2021/6628739.

[11] A. K. Pal and U. Kumar, "Retinal Blood Vessel Detection and Segmentation for Diabetic Retinopathy," vol. 3, no. 2, pp. 115–129, 2015.

[12] S. Choudhuri *et al.*, "Recent Advancements in Computational Drug Design Algorithms through Machine Learning and Optimization," *Kinases and Phosphatases*, vol. 1, no. 2, pp. 117–140, 2023, doi: 10.3390/kinasesphosphatases1020008.

[13] A. I. Journal, D. G. V, and J. J. T. S, "Improvement of osprey optimization algorithm and weighted RBM features for predicting the surface roughness and flank wear measurement using 1DCNN with ridge regression," *Mach. Sci. Technol.*, vol. 0, no. 0, pp. 1–50, 2025, doi: 10.1080/10910344.2025.2458050.

[14] N. Rica *et al.*, "Meat and Human Health—Current Knowledge and Research Gaps," pp. 1–17, 2021.

[15] V. D. Ganesh and R. Bommi, "FlankPix: An Image Segmentation Algorithm for Flank Wear Analysis in Monel K500 Turning," *Trait. du Signal*, vol. 41, no. 3, pp. 1625–1631, 2024, doi: 10.18280/ts.410349.