International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 1 (2026)



ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 1, 2026
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research

SECURE AND INTELLIGENT EARLY DISEASE DETECTION FRAMEWORK BASED ON ARTIFICIAL INTELLIGENCE AND BLOCKCHAIN

DR. P RAMESH BABU[1], ERRA SANJANA[2], DATTU DESHMUKH[3], BOMMAKANTI UDAY[4], BADDAM SUSHANTH REDDY[5]

PROFESSOR[1], UG SCHOLAR[2,3,4&5]

DEPARTMENT OF CSE, NARSIMHA REDDY ENGINEERING COLLEGE (UGC- AUTONOMOUS) MAISAMMAGUDA (V), KOMPALLY, SECUNDERABAD, TELANGANA-500100

PROBLEM STATEMENT

Early detection of human diseases is critical for improving treatment outcomes and reducing healthcare costs. However, traditional healthcare systems often face challenges such as delayed diagnosis, fragmented medical records, lack of secure data sharing, and limited access to intelligent decision-support systems. Medical data is usually stored in centralized databases, which increases the risk of data breaches, unauthorized access, and data manipulation. Artificial Intelligence (AI) has shown great potential in analyzing large medical datasets and identifying disease patterns at early stages, but concerns regarding data privacy, transparency, and secure data management still remain. On the other hand, Blockchain technology provides a decentralized and tamper-proof mechanism for storing and sharing healthcare data securely. Therefore, there is a need for an integrated framework that combines Artificial Intelligence for intelligent disease prediction and Blockchain technology for secure and transparent medical data management to enable reliable, early detection of human diseases while ensuring patient privacy and data integrity.

OBJECTIVES

Primary Objective

To develop a secure and intelligent framework for early detection of human diseases using Artificial Intelligence and Blockchain technology.

Specific Objectives

- To design an AI-based model capable of analyzing patient medical data for early disease prediction.
- To integrate Blockchain technology for secure storage and sharing of healthcare records.
- To ensure patient data privacy, integrity, and transparency using decentralized blockchain mechanisms.
- To develop a system that supports real-time disease monitoring and early diagnosis.
- To improve the accuracy and reliability of disease prediction using machine learning algorithms.
- To create a secure healthcare ecosystem that enables trusted collaboration between patients, doctors, and healthcare institutions.

ABSTRACT

Early detection of diseases plays a crucial role in improving healthcare outcomes and reducing mortality rates. With the rapid growth of healthcare data, traditional diagnostic systems often struggle to efficiently analyze large volumes of medical information while ensuring data security and privacy. This research proposes a Secure and Intelligent Early Disease Detection Framework based on Artificial Intelligence (AI) and Blockchain technology. The proposed system utilizes machine learning algorithms to analyze patient health data, medical records, and diagnostic parameters to predict diseases at an early stage. Artificial Intelligence enables the system to identify hidden patterns in medical datasets and provide accurate predictions that assist healthcare professionals in making informed decisions. To address concerns related to data privacy and security, Blockchain technology is integrated into the framework to provide decentralized, immutable, and transparent storage of medical records. Blockchain ensures secure sharing of healthcare data among authorized entities while preventing unauthorized access and data tampering. The combination of AI and Blockchain creates a reliable and secure healthcare infrastructure that enhances disease prediction accuracy, ensures patient data protection, and promotes trust in digital healthcare systems. The proposed framework aims to support healthcare providers in delivering timely diagnosis and improving overall patient care through intelligent and secure medical data management.

INTRODUCTION

THE process of screening, diagnosing, and subsequently following the development of a sickness or condition and the efficacy of a treatment is known as disease detection and monitoring [1], [2], [3], [4]. This methodology applies to a broad spectrum of illnesses requiring ongoing medical attention, with a predominant focus on chronic ailments such as heart problems, kidney disease, and liver disease. Fig. 1 illustrates the five major body parts, and this article has comprehensively reviewed diagnosing diseases related to these organs. The human heart, kidneys, liver, and thyroid constitute crucial components of the human anatomy. The heart's primary function involves pumping blood throughout the body and sustaining life through rhythmic beats. The circulated blood carries essential substances, including sugar and oxygen, vital for the proper functioning of the entire body. Maintaining a balanced sugar level in the bloodstream is imperative, as imbalances can have severe consequences, potentially leading to fatality. Additionally, kidneys filter and purify the blood, while the liver aids in various metabolic processes. Another vital organ, the thyroid, releases hormones

that regulate human metabolism, exerting significant control over various bodily functions. Ensuring the well being of these organs is paramount for sustaining a healthy life. Despite their significance, diseases affecting these organs are widespread globally. Particularly, heart disease and diabetes rank at the forefront cause of human mortality [5], [6]. Chronic kidney disease (CKD) is impacting over 10% of the global population, exceeding 800 million individuals [7]. Liver disease results in more than two million fatalities each year. It constitutes 4% of the total global mortality, translating to one out of every 25 deaths, with one-third of liver-related deaths occurring among females [8]. The prevalence of thyroid disease within the general population gradually increases yearly and is in the range of 0.3%–3.7% in the USA and 0.2%–3.5% in Europe [3]. These diseases take many lives every day [9], [10], [11], [12]. Concerns among clinicians are growing due to the increasing prevalence, and many researchers are actively working on using different artificial intelligence (AI) methodologies to alleviate their stress. Therefore, this study aims to understand the current trend, contributions, and applications for the early detection of these critical diseases, considering that all of them have highly complicated functions in the human body. For several reasons, early disease detection is essential. Early detection enables immediate care, raising the likelihood of successful remedy and cure [17], [18], [19]. When diagnosed early, diseases, including kidney and cardiovascular diseases (CVDs), have better prognoses. Accordingly, early detection enhances prognosis by allowing medical experts to implement suitable treatment plans and stop the disease from advancing [20]. This raises the patient's chances of survival and living standard. Additionally, premature detection lowers the risk of severe health problems by preventing difficulties linked to some disorders. Treating diseases in their early stages, typically less expensive than advanced and complex phases, also results in cost savings [21]. Early disease identification also has significant public health advantages, enabling the quick application of vaccinations and the management of infectious diseases. Finally, early detection encourages people to take charge of their health, adopt preventative measures, and make wise lifestyle decisions, ultimately boosting their overall health and lifespan. Many researchers are working on the diseases so that they can detect them as early as possible and provide better treatment as well. Fig. 2 shows the last five years' trend of publications for several diseases. It shows that the number of research works is increasing with time. Moreover, reduced death rates can be achieved by accurately and promptly detecting certain disorders. The heart is one of the necessary organs of the human body. Therefore, many research works have been incorporated regarding several CVD detection [22], [23], [24]. The growth for the detection of this area is incorporated in Fig. 3. The graph shows that work is growing in different domains. Accordingly, most families in the world have at least one diabetes patient, and its severity is increasing rapidly day by day. Researchers are also focusing on this disease detection and trying to reduce the harm it may cause. Fig. 4 conveys the growing rate of publications over the years in machine learning (ML), deep learning (DL), ensemble learning (EL), and the use of Blockchain technology is also rising. On the other hand, the kidneys, liver, and thyroid are also vital organs of the human body. However, researchers are not still paying attention to these diseases [25], [26], [27], [28]. Hence, complications are growing around the world. Figs. 5–7 illustrate that researchers are working on these diseases, and the number of publications is also growing, but they are not compatible with the growing rate of heart diseases and diabetes. Accordingly, the graphs show that the utilization of blockchain technology is very poor for these diseases. Blockchain technology has many benefits for storing patients' data in healthcare [1], [29], [30]. Key benefits include improving data security, fostering interoperability, and protecting

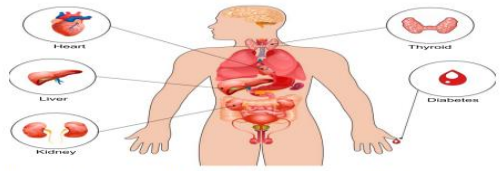


Fig. 1. Major organs related to the diagnosis of several diseases.

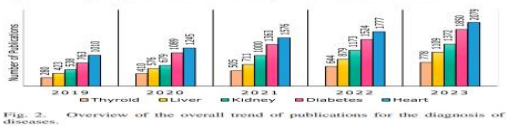


Fig. 2. Overview of the overall trend of publications for the diagnosis of diseases.

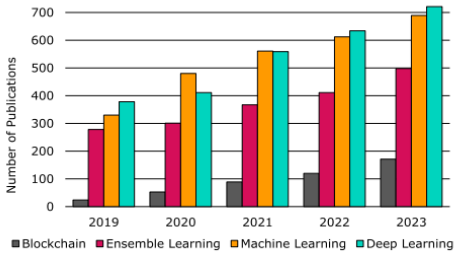


Fig. 3. Publication trend for heart disease prediction.

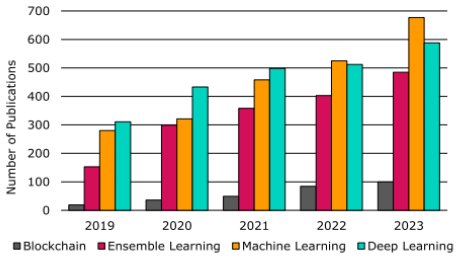


Fig. 4. Publication trend for diabetes prediction.

2.LITERATURE REVIEW

Early detection of human diseases is essential for improving patient outcomes, reducing healthcare costs, and enabling timely treatment. Recent advancements in **Artificial Intelligence (AI)** and **Blockchain technology** have significantly contributed to the development of intelligent and secure healthcare systems. AI provides powerful analytical capabilities for disease prediction, while blockchain ensures secure storage and sharing of medical data. This literature review summarizes the major research contributions related to AI-based disease detection and blockchain-enabled healthcare security.

Artificial Intelligence in Early Disease Detection

Artificial Intelligence has become a key technology in healthcare diagnostics due to its ability to analyze large medical datasets and identify hidden patterns. AI techniques such as **machine learning (ML), deep learning (DL), and neural networks** are widely used for detecting diseases at early stages.

A comprehensive survey by Kumar et al. reported that AI techniques have been successfully applied in diagnosing diseases such as **cancer, diabetes, Alzheimer's disease, cardiovascular disorders, and liver diseases** by analyzing medical images, electronic health records, and patient symptoms. These AI models

improve diagnostic accuracy and reduce the time required for medical analysis.

Recent studies also highlight that AI can process complex healthcare data and support personalized medicine, disease prediction, and early detection of health conditions. The ability of AI to identify patterns in large datasets has significantly improved disease identification and clinical decision-making.

Researchers have implemented algorithms such as **Random Forest, Logistic Regression, Support Vector Machine, and Gradient Boosting** for predicting diseases based on patient symptoms and historical data, demonstrating promising accuracy in early diagnosis.

Blockchain Technology in Healthcare Systems

Blockchain technology has emerged as an effective solution for addressing **data security, privacy, and transparency issues in healthcare systems**. It provides a decentralized and tamper-proof environment where patient records can be securely stored and shared among healthcare providers.

Studies indicate that blockchain enables secure exchange of medical data between hospitals, laboratories, and healthcare professionals while maintaining data integrity and patient privacy. This capability is particularly important for AI-based healthcare systems that rely on large datasets for training predictive models.

Research by Tagde et al. demonstrated that blockchain networks can support AI models by providing a reliable infrastructure for **data sharing, authentication, and authorization**, enabling healthcare professionals to access patient information securely.

Furthermore, blockchain-based healthcare frameworks improve **traceability, interoperability, and trust** among different healthcare stakeholders, making them suitable for distributed medical data management.

Integration of Artificial Intelligence and Blockchain for Disease Detection

Recent studies emphasize the combined use of **AI and blockchain** to develop secure and intelligent healthcare systems. The integration allows AI algorithms to analyze patient data while blockchain ensures data privacy and integrity.

A survey reviewing more than 100 research papers highlighted that combining AI with blockchain technologies can significantly enhance predictive health analytics and enable secure medical data sharing across institutions.

Similarly, advanced frameworks have been proposed where blockchain supports decentralized data management, while deep learning models analyze the stored data to detect diseases such as cardiovascular disorders, diabetes, and neurological diseases.

Research also shows that blockchain-enabled deep learning systems can improve diagnostic accuracy while protecting patient information through secure and tamper-proof data storage.

Explainable AI and Intelligent Decision Support in Healthcare

To enhance trust in AI-based medical systems, researchers are increasingly adopting **Explainable Artificial Intelligence (XAI)** techniques. XAI methods help healthcare professionals understand how AI models make predictions, which is crucial for clinical decision-making.

Recent studies propose AI-based disease diagnosis systems where prediction results and treatment recommendations are stored on a **permissioned blockchain network**, allowing medical experts to verify and validate the decisions collaboratively.

Such systems combine AI-driven analytics with blockchain-based validation mechanisms, enabling secure and transparent healthcare decision support systems.

Challenges in AI-Blockchain Healthcare Systems

Despite the promising advantages of AI and blockchain integration, several challenges remain:

- **Data privacy and regulatory compliance** in healthcare systems
- **High computational complexity** of deep learning models
- **Scalability issues in blockchain networks**
- **Limited availability of large, high-quality medical datasets**
- **Integration challenges between AI models and blockchain infrastructure**

Researchers also note that many blockchain-based healthcare systems remain at the **prototype or proof-of-concept stage**, requiring further real-world implementation and evaluation.

Research Gap and Future Directions

Although many studies have explored AI-based disease detection and blockchain-based data security independently, only a limited number of frameworks integrate both technologies effectively. Future research should focus on:

- Developing **scalable AI-blockchain healthcare architectures**
- Implementing **federated learning with blockchain for privacy-preserving AI models**
- Building **real-time disease prediction systems** using medical IoT devices
- Enhancing **explainability and transparency** in AI-based healthcare systems

Such improvements will enable the creation of **secure, intelligent, and reliable healthcare platforms** capable of detecting diseases at early stages.

Summary:
Existing research demonstrates that Artificial Intelligence plays a crucial role in analyzing healthcare data and detecting diseases at early stages, while blockchain technology ensures secure and transparent data management. The integration of these technologies has the potential to revolutionize healthcare by enabling **secure data sharing, intelligent disease prediction, and trustworthy medical decision support systems**. However, further research is required to address challenges related to scalability, privacy, and real-world implementation.

3.SYSTEM ANALYSIS

EXISTING SYSTEM

- A study used ML to identify CVD predictors [71]. Recursion enhanced RF with an improved linear model (RERF-ILM) detected cardiac disease more accurately. This study compared their suggested ML algorithm to SVM, NB, KNN, DT, etc., and artificial neural network (ANN). Their approach was the most accurate at 96.6%. Their methodology reduces diagnostic costs and time and improves treatment accuracy. Support vector classifier (SVC), KNN, logistic regression (LR), stochastic gradient descent (SGD), tree-based ensemble, and XGBoost had been used to detect and classify heart disease severity [57]. SMOTE, a model with HB, finds the optimal hyperparameters. Additional trees in their algorithm improved accuracy to 99.2% and 98.52% for Cleveland and Statlog datasets, respectively. Hence, Statlog and Cleveland's datasets were used most for heart disease identification [22], [57], [59].
- Another study used a similar dataset to forecast CVD and created a unique heart disease clinical decision support system (HDCDSS) [22]. Their heart disease prediction model (HDPM) surpassed all ML algorithms tested in this article. NB, LR, MLP, SVM, DT, and RF were investigated, with HDPM providing 95.90% and 98.40% accuracy for Statlog and Cleveland datasets. The clinical decision support system (CDSS) helped to diagnose cardiac problems prematurely. Density-based spatial clustering of applications with noise (DBSCAN) detected and removed outliers, a hybrid synthetic minority over-sampling technique-edited nearest neighbor (SMOTE-ENN) balanced the experimental data administration, and XGBoost predicts CVD. UCI repository is a popular dataset used by various researchers for CVD. An article used this resource to diagnose heart illness and test their IHDSSO model with RF with 98.38% accuracy [72].
- Another study proposed a unique MLP for enhanced Brownian motion based on the dragonfly algorithm (MLP-EBMDA) method for predicting heart disease [59]. Unsupervised feature extraction techniques were implemented for disease prediction. The method's accuracy was 94.28%, which was pretty acceptable. In a study, Ali et al. [33] used ML-based techniques with a different dataset. RF and KNN approaches obtained the most prominent outcomes compared to other algorithms, such as LR, MLP, AdaBoost, and DT. They demonstrated that a basic ML approach can produce great results that can be highly beneficial for disease detection.
- Accordingly, Bahani et al. [73] designed a novel data-driven approach for building a fuzzy rule-based classification system for heart disease detection based on fuzzy clustering and language modifiers. The proposed system included a knowledge base that could be used to interpret the system and clarify the decision-making process. Three datasets were used to evaluate the recommended model. The article split three datasets into two to increase the number of examples. Hence, the accuracy of their proposed Fuzzy classification Rules Learning via Clustering (FCRLC) model was 83.17% for the Cleveland data and 80.46% for the CombinedHunVa data. Hyperparameter optimization (HPO) tuning was presented as an alternative method for predicting CVD [32]. Three HPO approaches, grid search, randomized search, and genetic programming (TPOT Classifier), were proposed to improve the performance of the RF classifier and XG Boost classifier model. Using the Z-Alizadeh Sani dataset, the RF with randomized search performed best, detecting the stenos of three vessels—LAD, LCX, and RCA—with accuracy rates of 80.2%, 73.6%, and 76.4%, respectively. By merging MOPSO with RF, a new method for predicting heart disease was developed [74].
- This study utilized six distinct datasets to detect cardiac disease, with the Statlog dataset yielding the highest accuracy at 88.26%. In another research, titled GAPSO-RF, a hybrid GA and PSO optimized approach based on RF was created and applied to identify the ideal characteristics that enhance the validity of prediction of heart disease [9]. The proposed technique facilitated the implementation of multivariate statistical analysis to choose the most prominent features from the starting populace. Unfortunately, the computational cost of this approach was extremely high, and there was a lack of assessment of other algorithms in this study, which needs to be completed to compare the suggested algorithm with other algorithms appropriately.

DISADVANTAGES

- The complexity of data: Most of the existing machine learning models must be able to accurately interpret large and complex datasets for Early Detection of Human Diseases.
- Data availability: Most machine learning models require large amounts of data to create accurate predictions. If data is unavailable in sufficient quantities, then model accuracy may suffer.
- Incorrect labeling: The existing machine learning models are only as accurate as the data trained using the input dataset. If the data has been incorrectly labeled, the model cannot make accurate predictions.

PROPOSED SYSTEM

We have developed a solid understanding of the use of these techniques in the diagnosis of chronic diseases such as CVD, diabetes, renal disease, liver disease, and thyroid disease by thoroughly reviewing the current research. Our research on using blockchain technology in healthcare has also revealed potential advantages, such as improved patient privacy, interoperability, and data security. We examined the existing literature and case studies to understand how blockchain can change patient data storage, monitoring, and sharing and provide insights into its usage in disease detection and monitoring.

In this article, we have provided concise and clear demonstrations of widely used algorithms to make them understandable to various audiences. With the help of these illustrations, people can better understand how various ML and DL algorithms are used to identify diseases and how knowledge is disseminated. After that, the generalization of the introduced framework based on the learnings from the literature review is one of the significant contributions of this study. The framework provides a holistic view of the disease detection process and utilizes ML, DL, and EL algorithms. We offer professionals and scholars a helpful road

map to negotiate the difficulties of disease detection and make wise algorithmic decisions by introducing this framework. Finally, this article has also introduced blockchain technology for future endeavors by those who want to implement this technology and ensure a secure system. The framework for blockchain technology will help aspirants to make decisions and give them a clear path to this process.

ADVANTAGES

The contribution of ML, DL, and EL is inevitable for diagnosing several chronic diseases. Consequently, this article prioritizes these domains for diagnosing CVD, diabetes, kidney,liver, and thyroid diseases. Many researchers have applied ML techniques to implement CVDs, diabetes, kidney, liver, and thyroid diseases. Generally, the implemented algorithms in ML are random forest (RF), support vector machine (SVM), K-means clustering (KMC), multilayer perceptron (MLP), Naive Bayes (NB), K-nearest neighbor (KNN), decision tree (DT), and so on [34], [57], [58]. Among them, RF, KMC, NB, and KNN algorithms are used widely due to their performance and popularity in various disease diagnoses. The proposed system incorporates the most popular and most used ML approaches as well as their architectures and advantages and disadvantages of these algorithms. RF is widely utilized in liver disease prediction. On the other hand, the application of KNN is more popular in disease detection for diabetes, kidney, liver, and thyroid diseases prediction. Additionally, Naive Bayes is applied to diagnose several heart problems. The RF approach is a well-known supervised ML technique for solving regressionand classification problems. It is generally acknowledged that a forest comprises numerous trees, and the more trees it contains, the more resilient it will be.

IMPLEMENTATION

ALGORITHMS

DECISION TREE CLASSIFIERS

Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decision making knowledge from the supplied data. Decision tree can be generated from training sets. The procedure for such generation based on the set of objects (S), each belonging to one of the classes C1, C2, …, Ck is as follows:

Step 1. If all the objects in S belong to the same class, for example Ci, the decision tree for S consists of a leaf labeled with this class Step 2. Otherwise, let T be some test with possible outcomes O1, O2,…, On. Each object in S has one outcome for T so the test partitions S into subsets S1, S2,… Sn where each object in Si has outcome Oi for T. T becomes the root of the decision tree and for each outcome Oi we build a subsidiary decision tree by invoking the same procedure recursively on the set Si.

GRADIENT BOOSTING

Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees.[1][2] When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms random forest.A gradient-boosted trees model is built in a stage-wise fashion as in other boosting methods, but it generalizes the other methods by allowing optimization of an arbitrary differentiable loss function.

K-NEAREST NEIGHBORS (KNN)

- Simple, but a very powerful classification algorithm
- Classifies based on a similarity measure
- Non-parametric
- Lazy learning
- Does not “learn” until the test example is given
- Whenever we have a new data to classify, we find its K-nearest neighbors from the training data

Example

- Training dataset consists of k-closest examples in feature space
- Feature space means, space with categorization variables (non-metric variables)
- Learning based on instances, and thus also works lazily because instance close to the input vector for test or prediction may take time to occur in the training dataset

LOGISTIC REGRESSION CLASSIFIERS

Logistic regression analysis studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name *logistic regression* is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name *multinomial logistic regression* is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar.

Logistic regression competes with discriminant analysis as a method for analyzing categorical-response variables. Many statisticians feel that logistic regression is more versatile and better suited for modeling most situations than is discriminant analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as discriminant analysis does.

This program computes binary logistic regression and multinomial logistic regression on both numeric and categorical

independent variables. It reports on the regression equation as well as the goodness of fit, odds ratios, confidence limits, likelihood, and deviance. It performs a comprehensive residual analysis including diagnostic residual reports and plots. It can perform an independent variable subset selection search, looking for the best regression model with the fewest independent variables. It provides confidence intervals on predicted values and provides ROC curves to help determine the best cutoff point for classification. It allows you to validate your results by automatically classifying rows that are not used during the analysis.

NAÏVE BAYES

The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature .Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervised learning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight an explanation based on the representation bias. The naive bayes classifier is a linear classifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vector machine). The difference lies on the method of estimating the parameters of the classifier (the learning bias).

While the Naive Bayes classifier is widely used in the research world, it is not widespread among practitioners which want to obtain usable results. On the one hand, the researchers found especially it is very easy to program and implement it, its parameters are easy to estimate, learning is very fast even on very large databases, its accuracy is reasonably good in comparison to the other approaches. On the other hand, the final users do not obtain a model easy to interpret and deploy, they does not understand the interest of such a technique.

Thus, we introduce in a new presentation of the results of the learning process. The classifier is easier to understand, and its deployment is also made easier. In the first part of this tutorial, we present some theoretical aspects of the naive bayes classifier. Then, we implement the approach on a dataset with Tanagra. We compare the obtained results (the parameters of the model) to those obtained with other linear approaches such as the logistic regression, the linear discriminant analysis and the linear SVM. We note that the results are highly consistent. This largely explains the good performance of the method in comparison to others. In the second part, we use various tools on the same dataset (Weka 3.6.0, R 2.9.2, Knime 2.1.1, Orange 2.0b and RapidMiner 4.6.0). We try above all to understand the obtained results.

RANDOM FOREST

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance.

The first algorithm for random decision forests was created in 1995 by Tin Kam Ho[1] using the random subspace method, which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg.

An extension of the algorithm was developed by Leo Breiman and Adele Cutler, who registered "Random Forests" as a trademark in 2006 (as of 2019, owned by Minitab, Inc.).The extension combines Breiman's "bagging" idea and random selection of features, introduced first by Ho[1] and later independently by Amit and Geman[13] in order to construct a collection of decision trees with controlled variance.

Random forests are frequently used as "blackbox" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

SVM

In classification tasks a discriminant machine learning technique aims at finding, based on an *independent and identically distributed* (*iid*) training dataset, a discriminant function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminant classification function takes a data point x and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminant approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the equation for a multidimensional surface that best separates the different classes in the feature space.

SVM is a discriminant technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyperplane parameter—in contrast to *genetic algorithms* (*GAs*) or *perceptrons*, both of which are widely used

for classification in machine learning. For perceptrons, solutions are highly dependent on the initialization and termination criteria. For a specific kernel that transforms the data from the input space to the feature space, training returns uniquely defined SVM model parameters for a given training set, whereas the perceptron and GA classifier models are different each time training is initialized. The aim of GAs and perceptrons is only to minimize error during training, which will translate into several hyperplanes' meeting this requirement.

MODULES

The proposed system is divided into several functional modules that ensure **accurate disease prediction, secure medical data handling, and transparent healthcare management** using Artificial Intelligence (AI) and Blockchain technology.

Patient Data Collection Module

This module collects health-related data from patients through various sources such as **electronic health records (EHR), wearable devices, IoT medical sensors, and mobile health applications**.

The collected data may include:

- Body temperature
- Heart rate
- Blood pressure
- Medical history
- Laboratory test reports
- Lifestyle information

This module ensures that all patient data is gathered in a structured and standardized format for further processing.

Data Preprocessing and Feature Extraction Module

Raw medical data may contain **missing values, noise, or redundant information**. This module performs preprocessing operations such as:

- Data cleaning
- Data normalization
- Handling missing values
- Feature selection and extraction

These steps improve the quality of data before feeding it into the AI models for accurate disease prediction.

Artificial Intelligence-Based Disease Prediction Module

This is the core analytical module of the system. AI and machine learning algorithms are used to analyze patient health data and detect possible diseases at an early stage.

Possible algorithms used include:

- Random Forest
- Support Vector Machine (SVM)
- Neural Networks
- Deep Learning models

The module identifies patterns in medical data and predicts the **probability of diseases such as diabetes, heart disease, or respiratory disorders**.

Blockchain-Based Data Security Module

Blockchain technology is used to ensure **secure storage and sharing of medical data**. In this module:

- Patient records are stored as **encrypted blocks**.
- Each transaction is recorded in a **distributed ledger**.
- Unauthorized access and data tampering are prevented.

This ensures **data integrity, transparency, and privacy** in healthcare data management.

Smart Contract Management Module

Smart contracts are automated programs stored on the blockchain that control **data access and sharing policies**.

Functions include:

- Granting secure access to doctors and healthcare providers
- Managing patient consent for data sharing
- Automating medical record verification

This module ensures that only authorized users can access sensitive medical information.

Alert and Notification Module

Once the AI model detects potential health risks, this module sends alerts to:

- Patients
- Doctors
- Healthcare providers

Notifications may include **early disease warnings, recommended medical tests, or preventive measures**, enabling timely medical intervention.

Visualization and Reporting Module

This module presents the analysis results in an easy-to-understand format through:

- Health dashboards
- Disease prediction reports
- Risk assessment charts

Doctors and healthcare professionals can use these reports to make **informed clinical decisions**.

System Monitoring and Management Module

This module monitors the overall functioning of the system and manages:

- User authentication
- System performance
- Data access logs
- Blockchain transactions

It ensures the system operates efficiently and securely.

CONCLUSION

Disease diagnosis and surveillance are essential for improving patient outcomes and public health. Chronic conditions such as CVDs, diabetes, renal diseases, liver diseases, and thyroid diseases could be identified early using AI technologies, significantly improving the patients' living standards and increasing the likelihood of effective treatment. Additionally, early identification of these deep-rooted diseases enables the execution of suitable treatment strategies to cease disease proliferation and lessen potential complications, reducing costs and optimizing healthcare resources. Combining ML, DL, and EL algorithms holds enormous promise for delivering precise and fast assessments, assisting healthcare professionals in making knowledgeable decisions, and providing patients with individualized care. Additionally, blockchain technology in healthcare can enhance patient data security, compatibility, and privacy, creating opportunities for more efficient and patient-centered healthcare systems. There is a lot of potential for revolutionizing disease management and eventually saving more lives globally as research into disease detection and technological developments continue to develop. Therefore, to address the rising incidence of chronic diseases and enhance global health outcomes, researchers and healthcare professionals should work together to explore creative solutions that harness the power of these technologies. This article has observed some research works to incorporate a comprehensive review and understand the diagnosis of some chronic diseases and the necessary points researchers must focus on. Furthermore, the fundamental architecture of well-known algorithms is showcased to give prospects a general understanding of the critical components of several well-known algorithms and their significance. In light of this, a generalized blockchain architecture has also been created as a paradigm for future endeavors. However, in the future, the algorithms' efficacy will be validated by appropriate results, highlighting the unique aspects of the recommended approach. To analyze the generalized framework in this article, blockchain technology—an innovative method in healthcare informatics—will also be considered when evaluating the model and protecting data. Accordingly, we are committed to enlarge the analyzing process to include more diseases in the future.To conclude, ML, DL, and EL improve the process of diagnosing a variety of diseases, and the precision of the systems that are created is evidence of this. To assess the recent advancements in healthcare informatics technology, this article concentrates on applying these technologies to several disorders.

REFERENCES

[1] M. Chen et al., "Blockchain-enabled healthcare system for detection of diabetes," J. Inf. Secur. Appl., vol. 58, May 2021, Art. no. 102771,
doi: 10.1016%2Fj.jisa.2021.102771.

[2] E. Elsedimy, S. M. AboHashish, and F. Algarni, "New cardiovascular disease prediction approach using support vector machine and quantum-behaved particle swarm optimization," Multimedia Tools Appl., vol. 83, pp. 1–28, Mar. 2024.

[3] A. Sutradhar et al., "An intelligent thyroid diagnosis system utilising multiple ensemble and explainable algorithms with medical supported
attributes," IEEE Trans. Artif. Intell., vol. 5, no. 6, pp. 2840–2855, Jun. 2024.

[4] F. Ramoliya, K. Darji, R. Gupta, R. Kakkar, S. Tanwar, and A. Kumari, "Dl-based secure wearable scheme for remote patient data monitoring in healthcare 4.0," in Proc. Int. Conf. Comput., Commun., Intell. Syst. (ICCCIS), Piscataway, NJ, USA: IEEE Press, 2023, pp. 757–762.

[5] Z. M. Younossi, G. Wong, Q. M. Anstee, and L. Henry, "The global burden of liver disease," Clin. Gastroenterol. Hepatol., vol. 21, no. 8,
pp. 1978–1991, 2023. [Online]. Available: https://www.sciencedirect.
com/science/article/pii/S1542356523003154

[6] H. Bhatt et al., "Artificial neural network-driven federated learning for heart stroke prediction in healthcare 4.0 underlying 5g," Concurrency Comput.: Pract. Exp., vol. 36, no. 3, 2024, Art. no. e7911.

[7] A. Jayaraman, P. K. Jangalapuram, C. Ambayiram, and M. J. K. Alphonse, "An optimal Bi-gated recurrent unit and ResNET using IoT for healthcare monitoring," Expert Syst. Appl., vol. 247, 2024, Art. no. 122818.

[8] B. Maya1 and R. Jürgen, "A public health perspective on mitigating the global burden of chronic liver disease," Hepatology, vol. 79, no. 2, pp. 451–459, 2024.

[9] M. G. El-Shafiey, A. Hagag, E.-S. A. El-Dahshan, and M. A. Ismail, "A hybrid GA and PSO optimized approach for heart-disease prediction
based on random forest," Multimedia Tools Appl., vol. 81, no. 13, pp. 18155–18179, Mar. 2022, doi: 10.1007%2Fs11042-022-12425-x.

[10] S. Padhy, S. Dash, S. Routray, S. Ahmad, J. Nazeer, and A. Alam, "IoTbased hybrid ensemble machine learning model for efficient diabetes mellitus prediction," Comput. Intell. Neurosci., vol. 2022, pp. 1–11, May 2022, doi: 10.1155%2F2022%2F2389636.

[11] S. M. M. Elkholy, A. Rezk, and A. A. E. F. Saleh, "Early prediction of chronic kidney disease using deep belief network," IEEE Access, vol. 9, pp. 135542–135549, 2021, doi: 10.1109%2Faccess.2021.3114306.

[12] S. Tanwar, A. Kumari, D. Vekaria, N. Kumar, and R. Sharma, "An AI-based disease detection and prevention scheme for COVID-19,"
Comput. Elect. Eng., vol. 103, 2022, Art. no. 108352.

[13] P. M. Bossuyt et al., "Toward complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative," Amer. J. Clin. Pathol., vol. 119, no. 1, pp. 18–22, 2003.
[14] M. Sguanci, S. Mancin, M. Piredda, and M. G. De Marinis, "Protocol for conducting a systematic review on diagnostic accuracy in clinical research," MethodsX, vol. 12, 2024, Art. no. 102569.
[15] D. F. Stroup et al., "Meta-analysis of observational studies in epidemiology: A proposal for reporting," JAMA, vol. 283, no. 15, pp. 2008– 2012,