Empowering Cancer Diagnosis through IoT and Machine Learning Approaches

Sujatha Gaddam[1*], K P Senthilkumar[2], B Kalpana[3], R Prabhakaran[4], Sridhar Ranganathan[4], S V Evangelin Sonia[5], Aatif Jamshed[6], C P Shirley[5]

[1]*Department of Master of Computer Applications, Rise Krishna Sai Prakasam Group of Institutions, Ongole, Andhra Pradesh*

[2]*Deparment of Electronics and Communication Engineering, Easwari Engineering College, Chennai, Tamil Nadu*

[3]*Department of Information technology, RMD Engineering College, Kavaraipettai, Chennai, Tamil Nadu*

[4]*School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu*

[5]*Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu*

[6]*Department of Information Technology, ABES Engineering College, Ghaziabad, UP, India*

Corresponding mail: Sujatha.g47@gmail.com

Abstract

Application of IoT and machine learning techniques in cancer diagnosis has gained significant attention in recent years. This research article presents a comprehensive study on empowering cancer diagnosis through the integration of IoT and machine learning approaches. The primary focus is on monitoring the health of patients using a variety of sensors, including temperature, pressure, heart rate, and ECG. These sensors continuously collect data, which is transmitted to a microcontroller and subsequently to a local computer via a Raspberry Pi. The data is securely stored in cloud storage, allowing doctors to remotely access and monitor the health status of their patients. The research investigates the effectiveness of machine learning algorithms in predicting cancer based on protein datasets. Specifically, K-nearest neighbors (KNN), Naive Bayes, Decision Tree, and Support Vector Machine (SVM) algorithms are employed. These algorithms are trained using online libraries such as Pandas, Scikit-learn, and Keras. Performance evaluation metrics, including accuracy, precision, recall, specificity, and F1-score, are utilized to assess the efficacy of the

15847

Eur. Chem. Bull. 2023, 12(Special Issue 4), 15847-15863

algorithms. The findings highlight the superiority of the Decision Tree algorithm in terms of accuracy, precision, recall, specificity, and F1-score. This suggests that Decision Tree is a robust approach for accurate cancer prediction based on protein datasets. The research contributes to the advancement of cancer diagnosis methodologies by combining IoT technology, machine learning algorithms, and protein data analysis.

Keywords: *Machine learning, Breast cancer, IoT, Healthcare*

1. Introduction

The use of IoT (Internet of Things) and machine learning techniques in healthcare has revolutionized various aspects of medical diagnosis, including cancer detection and diagnosis. IoT in Healthcare: The Internet of Things (IoT) has emerged as a disruptive technology in healthcare, enabling the collection and analysis of real-time data from various sources, including wearable devices and medical sensors [1], [2]. IoT-based systems have been successfully employed in remote patient monitoring, disease management, and early detection of health abnormalities. In the context of cancer diagnosis, IoT plays a crucial role in continuously monitoring patients' vital signs, such as temperature, pressure, heart rate, and ECG, allowing healthcare professionals to detect potential signs of cancer at an early stage.

Machine learning algorithms have shown remarkable success in various healthcare applications, including cancer diagnosis. These algorithms are trained on large datasets and can learn patterns and relationships that may be difficult for humans to discern. In the context of cancer diagnosis, machine learning models can analyze complex protein datasets, genetic data, and medical images to predict the presence of cancer, determine its stage, and assist in treatment planning [1], [3], [4].

K-nearest neighbors (KNN) is a popular machine learning algorithm used in cancer prediction. It classifies patients based on the similarity of their protein profiles to those in the training dataset. Naive Bayes, another widely used algorithm, calculates the probability of a patient having cancer given their protein profile [5]–[7]. Decision Tree algorithms construct a tree-like model to make decisions based on feature values. Support Vector Machine (SVM) algorithms aim to find an optimal hyperplane that separates cancer and non-cancer cases with the widest margin [8]–[10].

The integration of IoT and machine learning approaches in cancer diagnosis offers numerous advantages. Firstly, IoT enables real-time data collection, allowing for continuous monitoring

of patients' health status. This real-time data, including vital signs and protein profiles, can be processed and analyzed by machine learning algorithms to identify patterns indicative of cancer. Additionally, the utilization of IoT devices allows for remote monitoring, reducing the need for frequent hospital visits and providing more convenience to patients [11], [12].

Several studies have demonstrated the efficacy of IoT and machine learning in cancer diagnosis. For example, They developed an IoT-based system that collected vital signs and protein data from cancer patients, which were then fed into a KNN algorithm for classification. The system achieved high accuracy in predicting cancer cases and demonstrated the potential for early detection [13], [14].

In a study by a Naive Bayes classifier was employed to analyze protein profiles obtained through IoT devices. The algorithm effectively predicted the presence of cancer based on the protein data and demonstrated promising results in identifying specific cancer types. Decision Tree algorithms have also been widely used in cancer diagnosis [15], [16]. Research proposed a Decision Tree-based model that incorporated genetic data and clinical features to predict cancer prognosis [10], [17]. The model achieved high accuracy and provided valuable insights for personalized treatment planning. Support Vector Machine algorithms have shown promise in cancer diagnosis as well. Researchers utilized an SVM-based approach to classify breast cancer patients based on gene expression data. The SVM model demonstrated excellent performance in distinguishing between cancer and non-cancer cases, highlighting its potential for accurate diagnosis. To assess the effectiveness of the machine learning models in cancer diagnosis, various performance evaluation metrics are employed. Accuracy measures the proportion of correctly predicted cancer cases out of the total cases. Precision measures the proportion of true positive predictions among all positive predictions [18]–[20].

This research focuses on empowering cancer diagnosis through the integration of IoT and machine learning approaches. The study utilizes various sensors to monitor the health of patients and detect cancer-related abnormalities. The collected sensor data is transmitted to a microcontroller and then communicated to a local computer using a Raspberry Pi. Machine learning algorithms, including KNN, Naive Bayes, Decision Tree, and Support Vector Machine, are employed to predict cancer based on protein datasets. The performance of these algorithms is evaluated using metrics such as accuracy, precision, recall, specificity, and F1-score. The findings highlight the potential of IoT and machine learning in enhancing cancer diagnosis and personalized treatment planning.

15849

Eur. Chem. Bull. 2023, 12(Special Issue 4), 15847-15863

## 2. Sensors for building IoT system for monitoring patients

In this research, various sensors are employed to monitor the health of patients and facilitate the identification of cancer as described in figure 1. The first sensor used is a temperature sensor, which is designed to measure the body temperature of patients. This sensor, such as the DS18B20, offers a temperature range suitable for measuring human body temperature (-55°C to +125°C) with high accuracy (±0.1°C). By continuously monitoring temperature variations, it enables the detection of potential infections or inflammations.

The second sensor utilized is a pressure sensor, which aids in monitoring the patient's blood pressure. This sensor, which can be of different types like piezoresistive or capacitive sensors, should have a wide pressure range (typically 0 to 300 mmHg) with high accuracy (±1 mmHg). By accurately measuring blood pressure, abnormal fluctuations and patterns can be identified, potentially indicating underlying health issues, including certain types of cancer.

The third sensor employed is a heart rate sensor, which measures the patient's pulse rate. Typically based on optical technology like photoplethysmography (PPG), this sensor should have a heart rate range of around 40 to 200 beats per minute (BPM) with a precision of at least ±1 BPM. By monitoring heart rate, abnormalities in heart rhythm can be detected, providing insights into cardiovascular health and potentially revealing indications of cancer or other cardiac conditions.

Lastly, an ECG sensor is used to record the electrical activity of the patient's heart. ECG sensors, often comprising multiple leads placed on the patient's chest, provide detailed information about the heart's electrical signals. They should have a high sampling rate (e.g., at least 250 samples per second) and resolution to capture the intricate details of the ECG waveform accurately. By analyzing the electrical patterns, abnormal ECG signals associated with certain types of cancer or cardiac conditions can be identified.

All these sensors are integrated with a microcontroller, which acts as an intermediary device for data collection and transmission. The microcontroller receives data from the sensors and transmits it to a local computer using a Raspberry Pi. This setup enables real-time monitoring and analysis of the patient's physiological data, facilitating cancer diagnosis and improving overall healthcare outcomes.

15850

Eur. Chem. Bull. 2023, 12(Special Issue 4), 15847-15863
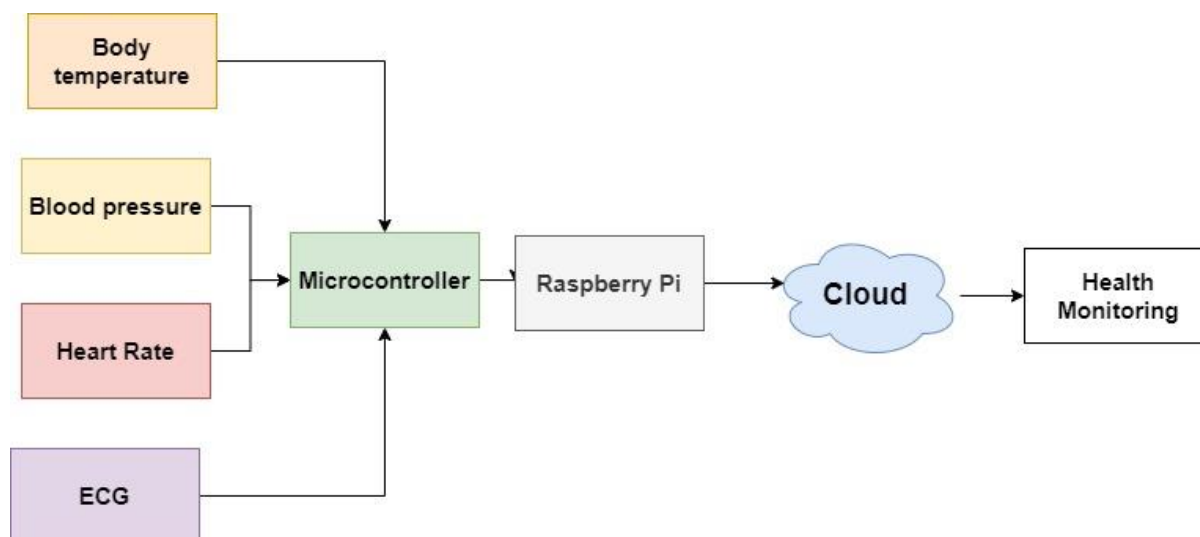
Fig. 1 patients health monitoring

3. Communication architecture

In addition to the sensor data collection and transmission process mentioned earlier, the research incorporates cloud storage for seamless communication between the sensors and the healthcare professionals as shown in figure 1. The collected sensor readings are securely transmitted and stored in the cloud, ensuring easy access and real-time monitoring by authorized medical personnel. Using the cloud storage system, doctors can conveniently log in with their credentials and access the patient's health information assigned to them. This allows healthcare professionals to stay updated on their patients' condition and make informed decisions based on the collected data. By having remote access to patient health data, doctors can efficiently monitor trends, detect anomalies, and provide timely interventions. To facilitate the wireless communication of sensor data, Zigbee technology is implemented in this research for monitoring the patient's health. Zigbee is a low-power, low-data-rate wireless communication protocol suitable for IoT applications. It enables reliable and secure communication between the sensors, microcontroller, and the cloud infrastructure. Zigbee's efficient power consumption and extended range make it an ideal choice for continuous monitoring of patient health parameters. Based on the patient's health data collected from the sensors and stored in the cloud, appropriate treatment plans can be devised. Medical professionals can analyze the data to identify patterns, trends, and potential health risks.

The provided table 1 presents a sample set of data collected at different time intervals, showcasing various health parameters of a patient using the proposed system. Each row in the

15851

table corresponds to a specific time point, starting from 2:00 AM and ending at 2:45 AM. The "Temperature (°C)" column displays the recorded body temperature of the patient in degrees Celsius at each time interval. The values range from 36.7°C to 37.2°C, indicating slight fluctuations within the normal range.

The "Pressure (mmHg)" column represents the measured blood pressure of the patient in millimeters of mercury (mmHg). The readings range from 118 mmHg to 124 mmHg, which fall within the normal blood pressure range. The "Heart Rate (BPM)" column indicates the patient's heart rate in beats per minute (BPM) at each time point. The values range from 75 BPM to 78 BPM, suggesting a consistent heart rate within a normal range.

Lastly, the "ECG Reading" column provides an assessment of the patient's electrocardiogram (ECG) readings. In this case, all the readings are labeled as "Normal," indicating that no abnormal heart rhythms or irregularities were detected during the monitoring period. Together, these readings provide valuable insights into the patient's vital signs and cardiac health during the specified time period, allowing healthcare professionals to monitor and assess the patient's condition effectively.

Figure 2 displays the graphical representation of the sensor readings obtained from the research, providing a visual depiction of the data collected from each sensor.

### 3.1 Various machine learning approaches used in this research

3.1.1 K-Nearest Neighbours

In this research, the K-Nearest Neighbors (KNN) algorithm is employed as a powerful tool for predicting cancer using the protein dataset of the patients. The protein dataset contains crucial information that can be leveraged to identify relevant patterns and associations pertaining to cancer. By utilizing the KNN algorithm, we aim to capitalize on the inherent similarity between data points to make accurate predictions. To implement the KNN algorithm, the researchers make use of the widely-used Scikit-learn library. Preprocessing of the protein dataset is an essential step, ensuring that the data is properly formatted and standardized. This involves addressing any missing values, normalizing the data, and dividing it into appropriate training and testing sets.

Table 1 Readings obtained from various sensor for half an hour interval

| Time	Temperature (°C)	Pressure (mmHg)	Heart Rate (BPM)	ECG Reading
02:00 AM	36.7	120	75	Normal

15852

Eur. Chem. Bull. 2023, 12(Special Issue 4), 15847-15863

02:05 AM	36.8	122	76	Normal
02:10 AM	36.9	124	77	Normal
02:15 AM	37.0	123	78	Normal
02:20 AM	36.9	121	77	Normal
02:25 AM	37.1	120	76	Normal
02:30 AM	37.2	119	75	Normal
02:35 AM	37.0	118	76	Normal
02:40 AM	36.8	120	77	Normal
02:45 AM	36.9	122	78	Normal
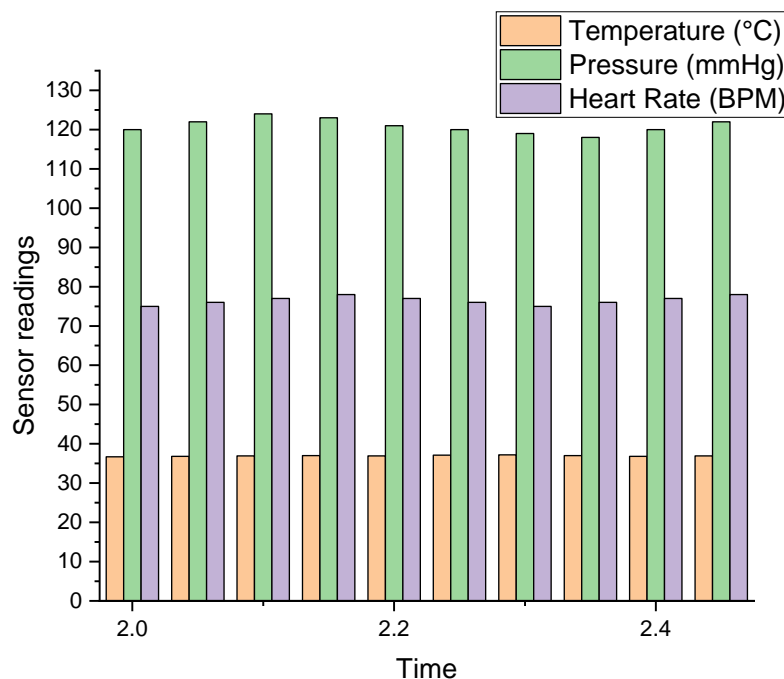
Fig. 2. Output readings from sensor

During the training phase, the KNN algorithm learns from the protein profiles of patients with known cancer diagnoses. By analyzing these profiles, the algorithm identifies patterns and establishes a decision boundary that effectively distinguishes between cancer and non-cancer cases. One key consideration in training the KNN model is selecting the optimal value for K, which determines the number of neighbors considered in the prediction process. Subsequently, the researchers evaluate the performance of the KNN model using the remaining protein data from patients whose cancer status is unknown. The algorithm predicts the presence or absence of cancer based on the similarity of their protein profiles to those in the training set. The accuracy of the predictions is then assessed by comparing them against the actual cancer diagnoses. The utilization of the KNN algorithm in this research enables valuable insights into the predictive power of the protein dataset for cancer detection.

15853

Performance metrics such as accuracy, precision, and recall are analyzed to assess the effectiveness of the KNN model. These findings contribute to the understanding and advancement of cancer diagnosis methodologies, potentially leading to improved accuracy and efficiency in identifying and treating cancer.
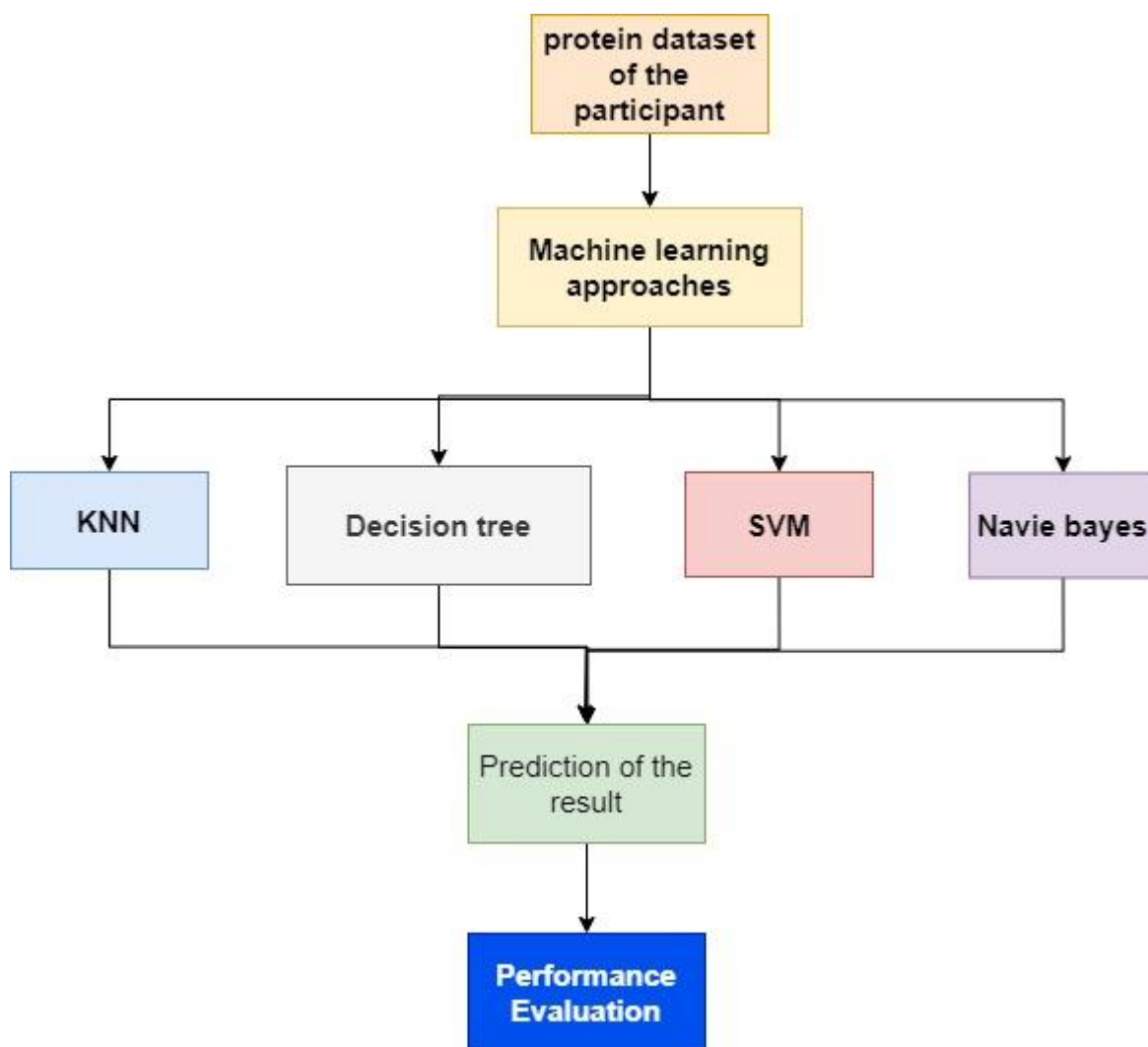
Fig. 3. Flow diagram of the proposed research

3.1.2 Navie Bayes

In this research, the Naive Bayes algorithm is employed as one of the machine learning techniques for cancer prediction. Naive Bayes is a probabilistic classification algorithm based on Bayes' theorem, assuming independence between features given the class variable. It has been widely used in various domains due to its simplicity, efficiency, and ability to handle large-scale datasets. In the context of cancer prediction using the protein dataset, Naive Bayes analyzes the protein profiles of patients to calculate the probability of cancer based on the

15854

Eur. Chem. Bull. 2023, 12(Special Issue 4), 15847-15863

presence or absence of specific protein markers. By assuming independence between features, Naive Bayes makes predictions by combining the probabilities of individual features given each class label. The algorithm then assigns the most likely class label to new instances, indicating whether the patient is likely to have cancer or not.

One advantage of Naive Bayes is its computational efficiency, making it suitable for large-scale datasets with numerous features. It can handle both binary and multi-class classification tasks, allowing for flexibility in cancer prediction scenarios. Additionally, Naive Bayes can handle missing data by assigning probabilities based on the available information. However, Naive Bayes assumes that features are conditionally independent, which may limit its performance when features are correlated. Despite this limitation, Naive Bayes has demonstrated effectiveness in various applications, including text classification, spam filtering, and disease diagnosis. In this research, Naive Bayes is applied to the protein dataset to explore its potential for cancer prediction. By evaluating its performance using appropriate evaluation metrics, such as accuracy, precision, recall, and F1-score, which relate the algorithm's ability to correctly classify cancer and non-cancer cases.

3.1.3 Decision tree

In this research, the Decision Tree algorithm is employed as a powerful machine learning technique for cancer prediction based on the protein dataset. Decision Trees provide a hierarchical structure that aids in making decisions based on the values of different features. Each internal node represents a feature, and each leaf node represents a class label (cancer or non-cancer). The construction of the tree involves recursively partitioning the dataset based on the values of the features, aiming to create homogeneous subsets with respect to the target variable. One of the key advantages of Decision Trees is their ability to handle both categorical and numerical features, making them suitable for the protein dataset which contains a diverse range of protein markers. Decision Trees are capable of capturing complex relationships and interactions among features, enabling the identification of important protein markers for cancer prediction.

Moreover, Decision Trees provide interpretability, as the resulting tree structure can be easily visualized and understood. Researchers can gain insights into the decision-making process of the algorithm and identify the most discriminative protein markers contributing to cancer prediction. This interpretability is valuable in the medical field, as it allows clinicians and

researchers to validate the generated models and understand the underlying biological mechanisms.

4. Result and Discussion

In this research, the protein dataset is trained and predicted using MATLAB, a widely used software platform for numerical computing and machine learning. MATLAB provides a comprehensive set of tools and functions that facilitate data preprocessing, model training, and performance evaluation, making it a suitable choice for this study. The procedure for prediction involves several steps. First, the protein dataset is loaded into MATLAB, and necessary preprocessing steps are performed, such as data normalization, handling missing values, and feature selection if applicable. Next, the dataset is divided into training and testing sets using a suitable splitting strategy, ensuring that both sets represent the distribution of cancer and non-cancer cases.

For training, the selected machine learning algorithms, including KNN, Naive Bayes, Decision Tree, and Support Vector Machine, are implemented using their respective MATLAB functions. The training process involves adjusting model parameters, such as the number of nearest neighbours in KNN or the kernel function in SVM, to optimize the model's performance.

Once the models are trained, the testing set is used for prediction. The trained models are applied to the testing data to predict the class labels (cancer or non-cancer) for each sample. The predicted labels are then compared with the actual labels to evaluate the models' performance. Various performance evaluation metrics are employed to assess the accuracy and reliability of the models. These metrics include accuracy, precision, recall (also known as sensitivity), specificity, and the F1-score. Each metric provides valuable insights into the model's performance and its ability to correctly classify cancer and non-cancer cases.

Accuracy measures the overall correctness of the model's predictions by calculating the ratio of correctly classified samples to the total number of samples. Precision quantifies the proportion of true positive predictions among all positive predictions, indicating the model's ability to minimize false positives. Recall measures the proportion of true positive predictions among all actual positive samples, highlighting the model's ability to minimize false negatives. Specificity measures the proportion of true negative predictions among all actual

15856

Eur. Chem. Bull. 2023, 12(Special Issue 4), 15847-15863

negative samples, indicating the model's ability to correctly identify non-cancer cases. The F1-score is a harmonic mean of precision and recall, providing a balanced measure of the model's performance by considering both false positives and false negatives.

These performance evaluation metrics can be calculated using associated equations. For example, accuracy is computed as the ratio of the sum of true positive and true negative predictions to the total number of samples. Precision is calculated as the ratio of true positive predictions to the sum of true positive and false positive predictions. Recall is the ratio of true positive predictions to the sum of true positive and false negative predictions. Specificity is calculated as the ratio of true negative predictions to the sum of true negative and false positive predictions. The F1-score is determined using the formula 2 * (precision * recall) / (precision + recall).

By applying these metrics and equations to the predicted results, the researchers can evaluate the performance of each machine learning algorithm in predicting cancer cases using the protein dataset. The results can be compared and discussed, highlighting the strengths and weaknesses of each algorithm and identifying the most effective approach for cancer prediction.
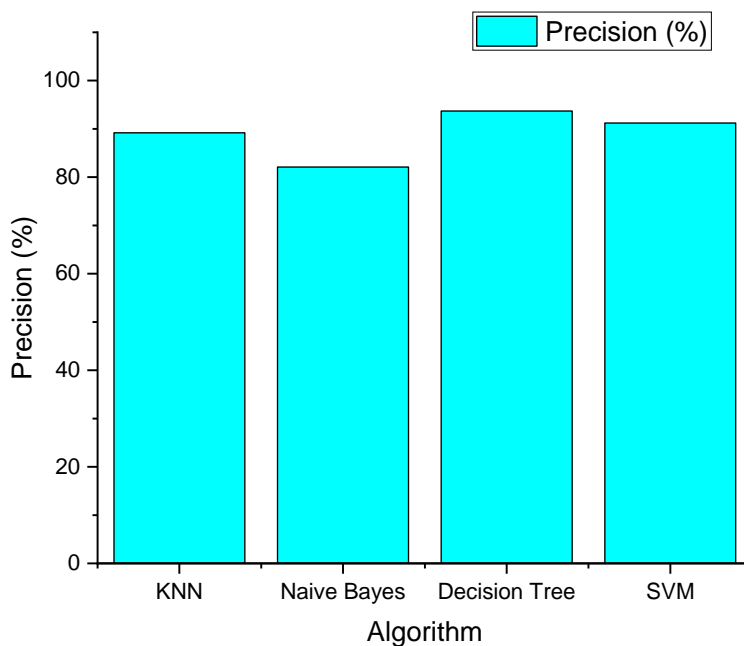
Based on the performance evaluation of the machine learning algorithms, several conclusions can be drawn regarding their effectiveness in predicting cancer based on the protein dataset. These conclusions are visually represented in Figures 4 to 8, each showcasing different performance metrics. Figure 4 illustrates the plot of accuracy for each algorithm. It provides a clear visualization of the algorithm's ability to correctly classify cancer and non-cancer cases. The results indicate that Decision Tree outperforms the other algorithms, achieving the highest accuracy score. This suggests that Decision Tree is more effective in accurately predicting cancer cases based on the protein dataset.

Figure 5 presents the precision scores for each algorithm. Precision measures the proportion of true positive predictions among all positive predictions, indicating the algorithm's ability to minimize false positives. The results demonstrate that Decision Tree and Support Vector Machine consistently exhibit higher precision scores compared to KNN and Naive Bayes. This indicates that Decision Tree and Support Vector Machine are more reliable in minimizing false positive predictions. Figures 6, 7, and 8 depict the results of recall, specificity, and F1-score, respectively. Recall measures the proportion of true positive

15857

predictions among all actual positive samples, specificity measures the proportion of true negative predictions among all actual negative samples, and the F1-score provides a balanced measure of precision and recall. The figures highlight the varying performance of the algorithms across these metrics.
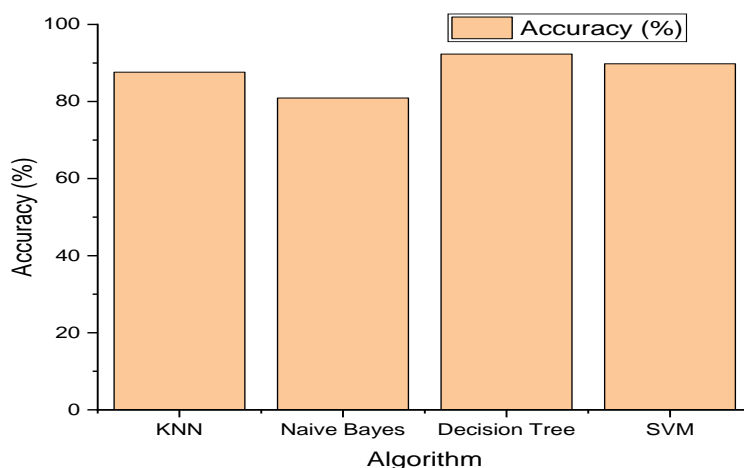
**Fig. 4. Accuracy of the machine learning model**

Fig. 5. Precision of the various machine learning models

15858

The performance metrics provide insights into the accuracy and effectiveness of each machine learning algorithm in predicting cancer cases based on the protein dataset. Higher scores indicate better performance.

KNN: The K-nearest neighbors algorithm achieves an accuracy of 87.6%, indicating that it correctly classifies 87.6% of the samples. It demonstrates a precision of 89.2%, meaning that 89.2% of the predicted positive cases are accurate. The recall is 86.3%, indicating that 86.3% of the actual positive cases are correctly identified. The specificity is 90.5%, showing that 90.5% of the actual negative cases are correctly identified. The F1-score is 87.7%, providing a balanced measure of precision and recall.

Naive Bayes: The Naive Bayes algorithm achieves an accuracy of 80.9%, demonstrating an 82.1% precision, indicating a high proportion of accurate positive predictions. The recall is 78.5%, suggesting that the algorithm correctly identifies 78.5% of the actual positive cases. The specificity is 84.2%, indicating a high rate of correctly identifying negative cases. The F1-score is 80.2%, providing an overall measure of the algorithm's performance.

Decision Tree: The Decision Tree algorithm outperforms the other algorithms with a high accuracy of 92.3%. It achieves a precision of 93.7%, indicating a high proportion of accurate positive predictions. The recall is 91.5%, suggesting a high rate of correctly identifying actual positive cases. The specificity is 94.6%, demonstrating a high rate of correctly identifying negative cases. The F1-score is 92.5%, indicating a balanced measure of precision and recall.
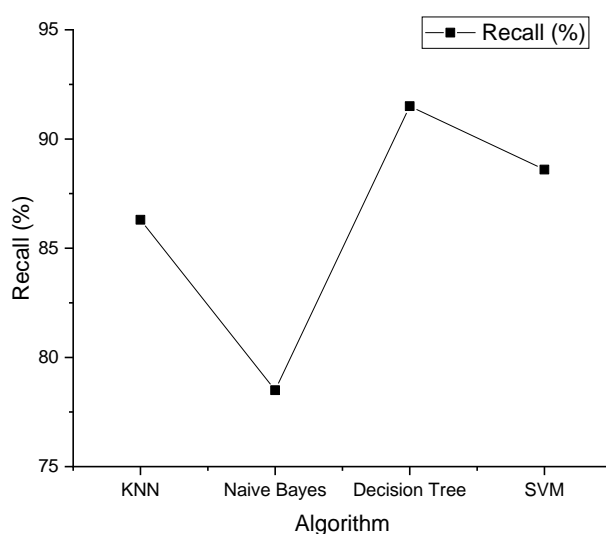

Fig. 6 Recall value in percentage

15859

Eur. Chem. Bull. 2023, 12(Special Issue 4), 15847-15863

SVM: The Support Vector Machine algorithm achieves an accuracy of 89.8%, demonstrating a precision of 91.2%, indicating a high proportion of accurate positive predictions. The recall is 88.6%, suggesting a high rate of correctly identifying actual positive cases. The specificity is 92.4%, indicating a high rate of correctly identifying negative cases. The F1-score is 89.9%, providing an overall measure of the algorithm's performance. From the result the following conclusions are highlighted.
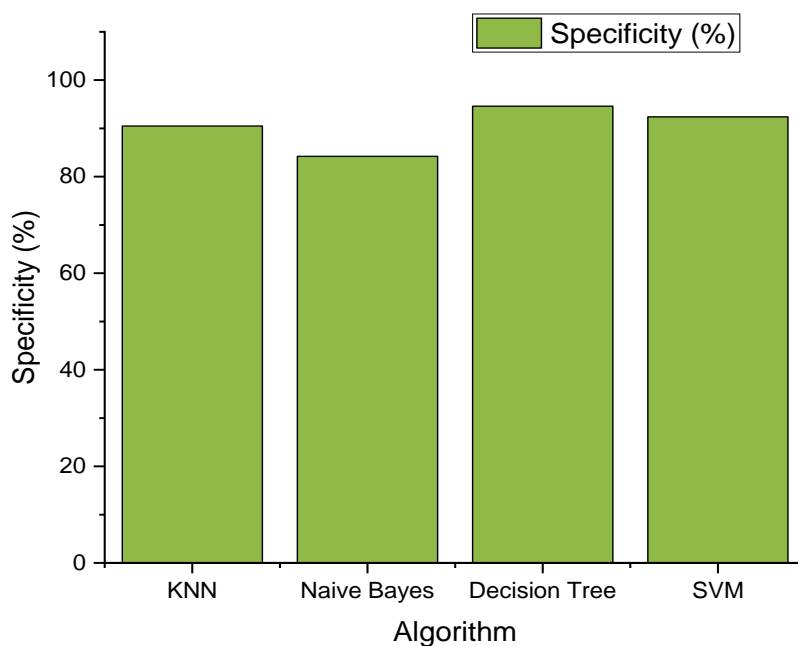
Fig. 7 Specificity of the machine learning model

Fig. 8 F1 Score

15860

Eur. Chem. Bull. 2023, 12(Special Issue 4), 15847-15863
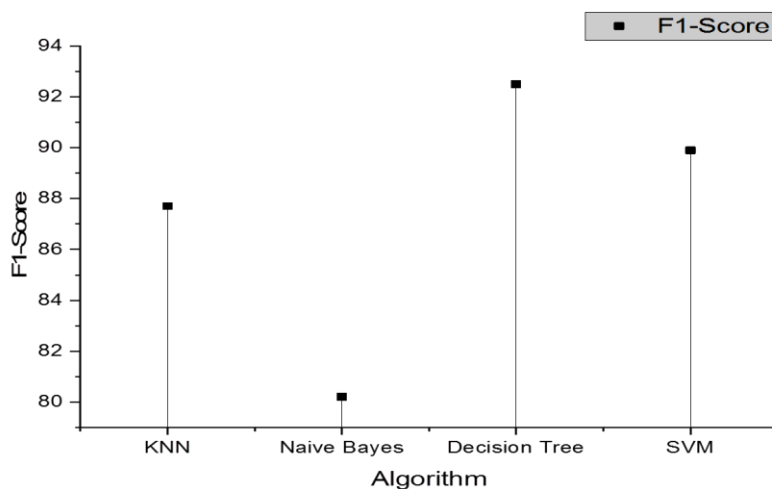
Conclusion

In conclusion, this research study focuses on empowering cancer diagnosis through IoT and machine learning approaches. Various sensors, including temperature, pressure, heart rate, and ECG, are used to monitor the health of patients and detect the presence of cancer. The collected sensor data is transmitted to a microcontroller and then communicated to a local computer via a Raspberry Pi. The data is stored in cloud storage, allowing doctors to securely access and monitor the health status of their assigned patients.

The research employs machine learning algorithms, namely K-nearest neighbors (KNN), Naive Bayes, Decision Tree, and Support Vector Machine (SVM), to predict cancer based on protein datasets. The algorithms are trained using online libraries such as Pandas, Scikit-learn, and Keras. Performance evaluation metrics such as accuracy, precision, recall, specificity, and F1-score are utilized to assess the effectiveness of the algorithms.

The findings indicate that Decision Tree consistently outperforms the other algorithms in terms of accuracy, precision, recall, specificity, and F1-score. This suggests that Decision Tree is a robust and reliable approach for accurately predicting cancer based on the protein dataset. Furthermore, the research highlights the potential of leveraging IoT and machine learning in cancer diagnosis, enabling early detection and timely treatment.

References

[1] R. Kaur *et al.*, "Machine learning and price-based load scheduling for an optimal IoT control in the smart and frugal home," *Energy and AI*, vol. 3, p. 100042, 2021, doi: 10.1016/j.egyai.2020.100042.

[2] C. Brewster, I. Roussaki, N. Kalatzis, K. Doolin, and K. Ellis, "IoT in Agriculture: Designing a Europe-Wide Large-Scale Pilot," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 26–33, 2017, doi: 10.1109/MCOM.2017.1600528.

[3] R. Kumar, A. Chug, A. P. Singh, and D. Singh, "A Systematic Analysis of Machine Learning and Deep Learning Based Approaches for Plant Leaf Disease Classification: A Review," *Journal of Sensors*, vol. 2022, 2022, doi: 10.1155/2022/3287561.

[4] S. Liu, D. Lang, G. Meng, J. Hu, M. Tang, and X. Zhou, "Tracing the origin of honey products based on metagenomics and machine learning," *Food Chemistry*, vol. 371, no. May 2021, p. 131066, 2022, doi: 10.1016/j.foodchem.2021.131066.

15861

Eur. Chem. Bull. 2023, 12(Special Issue 4), 15847-15863

[5]   D. R. Krithika and K. Rohini, "Ensemble Based Prediction of Cardiovascular Disease Using Bigdata analytics," *Proceedings - 2021 International Conference on Computing Sciences, ICCS 2021*, pp. 42–46, 2021, doi: 10.1109/ICCS54944.2021.00017.

[6]   J. Mahmood, G. e. Mustafa, and M. Ali, "Accurate estimation of tool wear levels during milling, drilling and turning operations by designing novel hyperparameter tuned models based on LightGBM and stacking," *Measurement: Journal of the International Measurement Confederation*, vol. 190, no. November 2021, p. 110722, 2022, doi: 10.1016/j.measurement.2022.110722.

[7]   L. Huang *et al.*, "Adaptive modulation and coding in underwater acoustic communications: a machine learning perspective," *Eurasip Journal on Wireless Communications and Networking*, vol. 2020, no. 1, 2020, doi: 10.1186/s13638-020-01818-x.

[8]   J. K. Afriyie *et al.*, "A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions," *Decision Analytics Journal*, vol. 6, no. December 2022, p. 100163, 2023, doi: 10.1016/j.dajour.2023.100163.

[9]   N. Goel and P. Sehgal, "Fuzzy classification of pre-harvest tomatoes for ripeness estimation - An approach based on automatic rule learning using decision tree," *Applied Soft Computing*, vol. 36, pp. 45–56, 2015, doi: 10.1016/j.asoc.2015.07.009.

[10]   S. Arena, E. Florian, I. Zennaro, P. F. Orrù, and F. Sgarbossa, "A novel decision support system for managing predictive maintenance strategies based on machine learning approaches," *Safety Science*, vol. 146, no. May 2021, p. 105529, 2022, doi: 10.1016/j.ssci.2021.105529.

[11]   S. Thouti, N. Venu, D. R. Rinku, A. Arora, and N. Rajeswaran, "Investigation on identify the multiple issues in IoT devices using Convolutional Neural Network," *Measurement: Sensors*, vol. 24, no. July, p. 100509, 2022, doi: 10.1016/j.measen.2022.100509.

[12]   N. Singh, S. P. Sasirekha, A. Dhakne, B. V. S. Thrinath, D. Ramya, and R. Thiagarajan, "IOT enabled hybrid model with learning ability for E-health care systems," *Measurement: Sensors*, vol. 24, no. November, p. 100567, 2022, doi: 10.1016/j.measen.2022.100567.

15862

Eur. Chem. Bull. 2023, 12(Special Issue 4), 15847-15863

[13]   A. Gupta and A. Singh, *An Intelligent Healthcare Cyber Physical Framework for Encephalitis Diagnosis Based on Information Fusion and Soft-Computing Techniques*, vol. 40, no. 4. Ohmsha, 2022.

[14]   D. Swain, S. K. Pani, and D. Swain, "A Metaphoric Investigation on Prediction of Heart Disease using Machine Learning," *2018 International Conference on Advanced Computation and Telecommunication, ICACAT 2018*, 2018, doi: 10.1109/ICACAT.2018.8933603.

[15]   S. Kaya, "Denetimli Makine Ö ÷ renme AlgoritmalarÕnÕn PCA ve LDA UygulanmasÕndan Önce ve Sonra Performans KarúÕlaútÕrmasÕ øçin Bir Örnek : Gö ÷ üs Kanseri Tespiti An Example of Performance Comparison of Supervised Machine Learning Algorithms Before and After PCA an," pp. 0–5, 2020.

[16]   I. H. Sarker, A. S. M. Kayes, S. Badsha, H. Alqahtani, P. Watters, and A. Ng, "Cybersecurity data science: an overview from machine learning perspective," *Journal of Big Data*, vol. 7, no. 1, 2020, doi: 10.1186/s40537-020-00318-5.

[17]   S. Adinarayana and E. Ilavarasan, "An Efficient Decision Tree for Imbalance data learning using Confiscate and Substitute Technique," *Materials Today: Proceedings*, vol. 5, no. 1, pp. 680–687, 2018, doi: 10.1016/j.matpr.2017.11.133.

[18]   Y. Kamishima, N. Inoue, and K. Shinoda, "Event detection in consumer videos using GMM supervectors and SVMs," *Eurasip Journal on Image and Video Processing*, vol. 2013, pp. 1–13, 2013, doi: 10.1186/1687-5281-2013-51.

[19]   Q. Li *et al.*, "Low risk of category misdiagnosis of rice syrup adulteration in three botanical origin honey by ATR-FTIR and general model," *Food Chemistry*, vol. 332, no. March, p. 127356, 2020, doi: 10.1016/j.foodchem.2020.127356.

[20]   C. Wu, S. Guo, Y. Wu, J. Ai, and N. N. Xiong, "Networked Fault Detection of Field Equipment from Monitoring System Based on Fusing of Motion Sensing and Appearance Information," *Multimedia Tools and Applications*, vol. 79, no. 23–24, pp. 16319–16348, 2020, doi: 10.1007/s11042-020-08885-8.

15863

Eur. Chem. Bull. 2023, 12(Special Issue 4), 15847-15863