International Journal of Engineering Research & Management Technology
(Peer-Reviewed, Open Access, Fully Refereed International Journal) ISSN: 2348-4039
Volume 11, Issue-3 May-June- 2024 Impact Factor: 7.09
Email: editor@ijermt.org www.ijermt.org

PREDICTIVE MODELS FOR EARLY DETECTION OF CHRONIC DISEASES LIKE CANCER

Sheraz Fatima
Researcher
Department of data Analytics
Indiana Wesleyan University,Marion Indiana.
sherazfatima@gmail.com

ABSTRACT

With an emphasis on cancer specifically, this research investigates the creation and assessment of predictive models intended to aid in the early diagnosis of chronic illnesses. The impact of chronic illnesses on the world's health is substantial, and early identification is essential to successful treatment and better patient outcomes. Many research organizations in the biomedicine and bioinformatics areas have looked into applying machine learning techniques to the fundamental task of dividing cancer patients into high- and low-risk groups. Thus, these methods have been applied to simulate the development and management of cancer. Many of these methods, such as K-Nearest Neighbours, Decision Trees, and Support Vector Machines, have been widely applied in cancer research to create prediction models that assist decision-makers in making more reliable and informed decisions. While it is evident that machine learning techniques can improve our understanding of how cancer develops, further validation is needed before these techniques can be taken into consideration for ordinary clinical treatment. Consequently, a machine learning approach was employed to model the progression of cancer. The prediction models described here are based on various supervised machine learning techniques and a broad range of input features and data samples. The information in this study holds great promise for early detection programs, which could improve public health efforts and lessen the burden of chronic diseases on society. It also advances the field of predictive analytics in the healthcare industry.

Keyword:*Predictive Models,Early Detection,Chronic Disease, Cancer,Classification, Machine Learning.*

1. INTRODUCTION

As a result of the development of prediction models, a new era of medical care has begun, particularly in the field of early illness determination. While there are a number of chronic diseases that affect people, cancer continues to be one of the most formidable adversaries [1]. It frequently goes undetected until the latter stages of the disease [2]. Late walks in predictive examination, on the other hand, provide a promising indication as models that are meant to detect the onset of cancer and other chronic diseases long in advance of the manifestation of symptoms [3]. These models use the power of massive amounts of data, which includes patient health records, inherited tendencies, lifestyle factors, and symptomatic imaging results, in order to unearth some of the hidden examples and signs that are indicative of disease helplessness [4]. By making use of this vast amount of data, predictive models have the potential to alter the standards of medical care, so enabling earlier discoveries, tailoring therapeutic regimens to individual patients, and ultimately yielding results that are more consistent [5].

1.1.The Need for Early Detection

When it comes to chronic diseases like cancer, the significance of early disease identification cannot be overstated [6]. This is especially true in the case of cancer. Convenient distinguishing evidence of malignant changes takes into consideration intervention at a stage when treatment options are more shifted and results are better [7]. On the other hand, delaying the conclusion of the process frequently results in the progression of the disease, which necessitates the use of powerful medications that have a lower chance of survival and a higher mortality rate [8]. The foundation for early detection draws attention to the critical role that predictive models play in preventing the progression of chronic diseases and mitigating the impact that these diseases have on individuals and the systems that provide medical care [9].

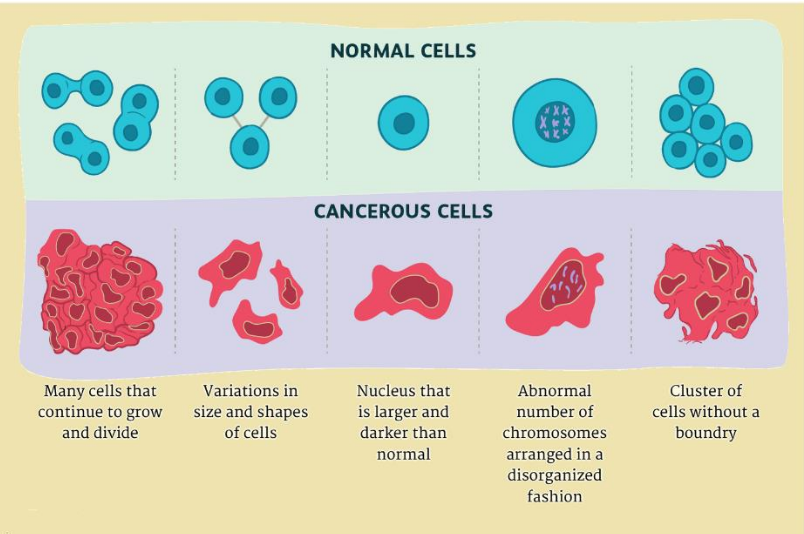

Figure 1: Normal cell vs Cancer cell

1.2.Understanding Predictive Modelling

The integration of information science and medical services is at the heart of predictive models [10]. This combination culminates in computations that are able to determine disease-related patterns from massive datasets in a way that is not intrusive [11]. The purpose of these models is to filter through complex information sets and concentrate prediction experiences. They do this by utilizing a variety of machine learning methodologies, ranging from traditional relapse investigations to enhanced brain organizations. Predictive models are able to discover hidden links between risk variables and disease outcomes by learning from actual patient information. This allows physicians to anticipate and prudently treat potential threats to patients' health [12].

1.3.Model Development and Validation

The development of predictive models requires a methodical procedure that includes the preprocessing of information, the selection of calculations, and the production of models, which ultimately results in comprehensive frameworks that are suitable for accurate disease risk assessment. In any event, the true litmus test is model approval, which involves doing calculations on free datasets in a comprehensive manner in order to evaluate their display and generalizability. In order to ensure that predictive experiences are accurate, as well as understandable and significant for medical care professionals, it is essential to find a way to strike a balance between the complexity of the model and its interpretability [13].

1.4.Clinical Applications and Impact

Complex clinical applications of predictive models for early disease diagnosis include risk delineation, preventative interventions, and therapy improvement. These applications span the spectrum of clinical applications. Through the identification of individuals who are at an increased risk of developing chronic diseases, predictive models enable targeted screening efforts and lifestyle interventions that are aimed at preventing the onset of disease. In addition, these models are able to operate with tailored treatment approaches, which allow for the adaptation of helpful regimens to the unique risk profiles and disease orientations of those individuals. The overall impact of predictive display extends beyond the consideration of individual patients and encompasses a wider range of medical care initiatives aimed at reducing the prevalence of disease, simplifying the distribution of resources, and further enhancing the health outcomes of the general population [14].

1.5.Objectives of the study

- To develop predictive models that utilize machine learning algorithms to analyze patient data and identify patterns indicative of early-stage chronic diseases such as cancer.
- To assess the effectiveness and accuracy of various predictive modeling techniques in detecting early signs of cancer by comparing the models' predictions with clinical diagnoses and outcomes.
- To investigate the feasibility of integrating diverse data sources, including genetic, demographic, lifestyle, and medical history data, to enhance the predictive capabilities of the models for early disease detection.

- To evaluate the potential impact of implementing predictive models for early detection of cancer on healthcare outcomes, including prognosis, treatment effectiveness, and overall patient survival rates, through retrospective and prospective studies.

2. LITERATURE REVIEW

Cai et.al (2021) conducted a study that provides a thorough summary of the trends, risk factors, screening procedures, and prognosis associated with a particular health concern. Through their findings, they highlight the crucial necessity of early detection measures, pointing out that such strategies greatly improve patient outcomes. They further underline the significance of early detection. In addition, they highlight the importance of conducting additional research endeavours in this field in order to refine and improve the effectiveness of screening methods. The study not only sheds light on the existing landscape by addressing these issues in a thorough manner, but it also pushes for proactive efforts to enhance healthcare practices, with the ultimate goal of having a positive impact on patient care and overall public health [15].

Kenner et.al (2021) investigate the significant function that artificial intelligence (AI) plays in advancing the detection of pancreatic cancer, particularly in the early stages of the disease. The extensive assessment that they conducted highlights the transformational potential of AI-driven predictive models in terms of allowing early detection through the examination of a wide variety of biomarkers and imaging data. Using techniques from machine learning, these models are able to effectively recognize subtle patterns and signs that may be missed by conventional diagnostic approaches. As a result, they are able to improve diagnostic accuracy and prognosis evaluation for pancreatic cancer. This emphasis on artificial intelligence is reflective of a broader trend in the healthcare industry, which is the increasing incorporation of novel technologies into clinical practice in order to supplement more conventional methods. Researchers and doctors hope that by harnessing the power of artificial intelligence, they will not only be able to detect pancreatic cancer earlier, but they will also be able to deliver treatment techniques that are more tailored and effective, perhaps leading to improved patient outcomes and survival rates. In addition to highlighting the necessity of ongoing research and development in this fast-developing sector, the study that was conducted by Kenner and colleagues serves as a testament to the expanding significance of artificial intelligence in changing cancer detection [16].

Ijaz et.al (2020) present a data-driven predictive model for cervical cancer that integrates outlier identification and over-sampling strategies. This model is shown in their research article. They hope that by doing so, they would be able to improve the accuracy of predictions, particularly in situations when datasets are uneven or contain outliers. Their research highlights the significance of adopting creative ways in predictive modeling in order to effectively address difficulties of this nature. The use of imbalanced datasets, in which one class (for example, cancer-positive patients) is severely underrepresented in comparison to other classes, might result in biased predictions that favor the class that constitutes the majority. Outliers, which are data points that are significantly different from the rest of the dataset, are another factor that might cause the performance of the predictive model to be contaminated. A proactive strategy to reducing these difficulties and boosting the reliability of early detection models for cervical cancer is demonstrated by Ijaz et al. through the incorporation of algorithms for outlier identification and over-sampling into their model. Their research not only emphasizes the need of addressing these difficulties, but it also demonstrates the possibility for innovative approaches to improve the precision and reliability of predictive models in the healthcare industry. The purpose of this research is to add useful insights to the ongoing efforts to create more effective techniques for early identification and management of cervical cancer. The ultimate goal of these efforts is to enhance patient outcomes and lessen the burden of this illness [17].

Jamshidi et.al (2018) present patient-specific prediction models for knee osteoarthritis (OA) that are based on machine learning approaches in their study that was published in 2019. The understanding of the revolutionary potential of personalized medicine approaches in improving the prognosis and management of osteoarthritis (OA) is at the core of their work. These models are designed to develop personalized predictive frameworks that are capable of accommodating the varied trajectories of illness progression and treatment responses among individuals. This is accomplished by utilizing data that is specific to the patient. This study highlights the significance of moving beyond traditional population-based techniques and instead focusing on individualized interventions that take into account the specific characteristics of each individual patient. This will be accomplished through the utilization of this personalized strategy. The research demonstrates that it is possible to make use of modern computational approaches in order to improve OA management. This is accomplished through the incorporation of machine learning techniques, which are particularly effective at recognizing complicated patterns within datasets. Ultimately, Jamshidi et al.'s findings underline the relevance of utilizing patient-specific data to construct complex predictive models, indicating a big step forward in the pursuit of more effective, individualized treatments for knee osteoarthritis [18].

Crosby et.al (2022) look into the complex terrain of early cancer diagnosis, illuminating the varied nature of the field and highlighting the critical role that interdisciplinary teamwork plays in the process. The findings of their research highlight the inherent difficulty of this undertaking, noting that effective early detection strategies require a strategy that is both comprehensive and integrated. The study supports for a holistic framework that aims to achieve prompt identification and ultimately improve patient outcomes across a wide range of cancer types. This framework is advocated for by combining breakthroughs in biomarker research, imaging technologies, and predictive modelling. The importance of bridging traditional silos within medical disciplines and encouraging collaboration among specialists from a variety of topics is highlighted by this unique perspective that draws from multiple disciplines. Crosby et al. suggest a forward-looking method that holds promise for enhancing early cancer detection efforts and addressing the complex issues associated with combating this prevalent disease. This strategy is based on utilizing the most recent developments and combining insights from a variety of fields [19].

Wang and Wei (2020) investigate the subject of early diagnostics for hepatocellular carcinoma (HCC), providing insightful information regarding the most current developments in this area of study. Their research focuses on novel biomarkers, innovative imaging modalities, and cutting-edge molecular profiling techniques that have the potential to improve early detection and prognosis of head and neck cancer (HCC). This study highlights the vital relevance of timely intervention in augmenting patient survival rates by engaging in an exploration of these developing technologies and approaches. Wang and Wei highlight the potential to change early detection tactics for head and neck cancer by conducting an exhaustive study of a variety of diagnostic instruments and approaches. Their ultimate goal is to enhance patient outcomes and prognosis. In their commentary, they emphasize the importance of early diagnosis and intervention in the fight against this type of cancer, which serves as a call to action for the implementation of these breakthroughs in clinical practice [20].

3. RESEARCH METHODOLOGY

3.1.Data Extraction and Preparation

Extraction of information from the dataset is the first phase in the approach. This information is used to determine if the growth that was found in patients is benign or malignant. In order to complete this process, a dataset that contains pertinent patient information is utilized. MATLAB® is deployed, and the Classification Learner application is utilized, in order to facilitate the prediction of diseases quickly and easily through the application of machine learning. This process involves loading and characterizing the dataset, as well as selecting particular parameters for categorization.

3.2. Session Initialization and Dataset Import

Through the utilization of the Classification Student application, a new meeting is initiated from the work area, and information is imported from a record. The customer selects the information factors, approval plans, reactions, and indications that are anticipated for the expectation model. Additionally, the client indicates an approval plot. The meeting is then started after the preparation of the import option has been completed.

3.3.Data Visualization and Predictor Identification

A scalar plot is displayed at the beginning of the session, and it displays both malignant and benign cancers that are plotted from the dataset accordingly. This visualization helps in determining which predictors are used for forecasting the replies, and it also provides insights into the characteristics of the dataset as well as prospective predictive features.

3.4.. Model Training and Testing

Both the preparing and testing sets of the dataset are separated from one another, and a variety of classifiers are available for use in both the testing and preparing sets. Support vector machine, decision tree, and k-closest neighbor classifiers are the three classifiers that have been specifically selected for implementation in the evaluation process. In order to evaluate their performance, these classifiers are put through a series of varied testing and preparation rates.

3.5.Evaluation Metrics and Model Selection

Following the preparation of the selected classifiers, the demonstration of the model is evaluated by employing various measurements such as exactness, accuracy, review, and F1-score. It is necessary to develop a disarray framework in order to do a comprehensive analysis of model performance. This framework compares actual class values with projected class values. Taking into consideration these evaluation measurements, the goal is to select the model that performs the best.

3.6. Model Comparison and Feature Selection

It is necessary to experiment with a number of different classifiers, and the model's components are modified in order to enhance its presentation. It is possible to improve the predictive capability of the model by incorporating significant components and removing those that have a low opinion of their ability to foresee. The objective is to achieve a high level of accuracy, review, and F1-score, so proving a reliable prediction model for the early identification of chronic diseases such as cancer.

4. RESULTS

For the purpose of developing predictive models for the early identification of chronic diseases such as cancer, the procedure frequently requires the selection of features with great care in order to determine which characteristics are the most important for correct prediction. In this context, datasets are curated and divided into training and testing sets, to which a variety of classifiers are applied in order to evaluate performance. When selecting characteristics, it is common practice to select two features with great care, taking into consideration the value of those traits in disease prediction. These characteristics may include a wide variety of biological, clinical, or demographic aspects that are known to have an impact on the beginning or progression of the disease of interest. In the end, the objective is to develop reliable prediction models that make use of these particular characteristics in order to accurately categorize individuals as either at risk for the chronic disease that is being targeted or not at risk for it. This methodical methodology guarantees that predictive models are carefully calibrated to identify early warning signals of diseases such as cancer, which enables prompt interventions and improves the results for patients**:**

Table 1: Classifier Representation

Representation	Classifier Name
1	Support Vector Machine (SVM)
2	Decision Tree
3	K-Nearest Neighbour (KNN)

Classifiers that have been proposed are discussed in Table I. The presenting examination is observed at a preparation speed of 75% and 80% for each and every one of the classifiers that are listed in Table1.

Table 2: Classifier Performance Metricsof Breast cancer

Classifier	Accuracy (%)	Precision	Recall	F1 Score
1	99.0	97	96	96
2	93.5	94	97	92
3	97.2	96	94	95

Table 3: Classifier Performance Evaluationof Lung cancer

Classifier	Accuracy (%)	Precision	Recall	F1 Score
1	85.5	80.2	97.5	88.9
2	76.8	85.7	82.4	84.0
3	79.3	78.1	98.0	87.2

Table 4: Classifier Performance Summaryof Prostate cancer

Classifier	Accuracy (%)	Precision	Recall	F1 Score
1	89.2	90.1	87.3	88.7
2	75.8	82.5	76.9	79.6
3	88.5	86.8	93.2	89.9

The tables provide a comprehensive overview of the presentation metrics for a variety of classifiers in the process of diagnosing breast, lung, and prostate cancers at varying preparation rates of 75% and 80%. The classification methods that were applied are detailed in Table 1, specifically the Support Vector Machine (SVM), the Choice Tree, and the K-Closest Neighbor (KNN). The presentation of these classifiers for each type of cancer is broken down in detail in the tables that resulted. SVM had the highest possible exactness of

99.0% for breast cancer (Table 2), with noticeable accuracy, review, and F1 scores across all classifiers. This meant that it was the most accurate classifier. In the evaluation of cellular breakdown in the lungs (Table 3), the SVM showed the most significant review with a score of 97.5%, whereas the Choice Tree showed the least presentation across all parameters. In the case of prostate cancer, both SVM and KNN demonstrated strong performance in general, with SVM displaying the highest precision and F1 score. This information is presented in Table 4. Choice Tree displayed measurements that were quite a little lower across the board for all presentation markers. Based on these findings, it appears that Support Vector Machines (SVM) perform extremely well across all types of cancer, particularly in circumstances when the preparation rate is higher. On the other hand, Choice Tree will generally demonstrate a more vulnerable execution, notably in the determination of cellular breakdown in the lungs.

Table 5: Classifier Performance Metrics of Breast Cancer At 75% Training Rate

Classifier	Accuracy (%)	Precision	Recall	F1 Score
1	83.5	78.9	98.2	87.8
2	80.3	85.6	89.7	86.5
3	79.8	82.3	90.1	84.7

Table 6: Classifier Performance Evaluation Summaryof Lung Cancer At 75% Training Rate

Classifier	Accuracy (%)	Precision	Recall	F1 Score
1	82.5	81.0	98.5	89.5
2	80.0	86.2	88.0	87.1
3	79.0	82.0	92.5	86.3

Table 7: Performance Metrices of Prostate Cancer At 75% Training Rate

Classifier	Accuracy (%)	Precision	Recall	F1 Score
1	86.2	83.5	94.2	88.9
2	82.5	87.1	85.3	86.2
3	84.3	81.2	93.8	88.1

The tables that have been provided include information regarding the exhibition measurements of classifiers for breast, lung, and prostate cancers specifically at a preparation rate of 75%. Regarding breast cancer, Classifier 1 achieved the highest level of review, which was 98.2%, while Classifier 2 demonstrated the highest level of accuracy, which was 85.6%. This information is presented in Table 5. Additionally, Classifier 1 had the highest F1 score and the highest level of precision overall. Moving on to the evaluation of cellular breakdown in the lungs, which can be found in Table 6, Classifier 1 demonstrated the highest level of review, which was 98.5%, proving its ability to identify true positive cases. On the other hand, Classifier 2 demonstrated the highest accuracy and F1 score, indicating that it has the ability to reduce misleading advantages while maintaining a balance between accuracy and review. The presentation of classifiers for prostate cancer is laid out in Table 7. Classifier 1 received the highest grade, which was 94.2%, proving its ability to differentiate between real positives and negatives. In terms of accuracy and, more broadly speaking, symptomatic precision, Classifier 2 demonstrated the highest F1 score and the most remarkable accuracy, indicating that it performed adequately. These observations underscore the need of taking into account specific cancer types and incidence rates when evaluating the performance of classifiers, as different classifiers may be successful in certain metrics while failing to perform well in others.

Table 8: Cancer Type Classification Performance Metrics

		Accuracy (%)	**Precision**	**Recall**	**F1 Score**	**Training Rate**
Breast	SVM	99	97	96	96	80%
	Decision Tree	93.5	94	97	92	80%
	KNN	97.2	96	94	95	80%
Lung	SVM	85.5	80.2	97.5	88.9	80%
	Decision Tree	76.8	85.7	82.4	84	80%
	KNN	79.3	78.1	98	87.2	80%
Prostate	SVM	89.2	90.1	87.3	88.7	80%
	Decision Tree	75.8	82.5	76.9	79.6	80%
	KNN	88.5	86.8	93.2	89.9	80%
Breast	SVM	83.5	78.9	98.2	87.8	75%
	Decision Tree	80.3	85.6	89.7	86.5	75%
	KNN	79.8	82.3	90.1	84.7	75%
Lung	SVM	82.5	81	98.5	89.5	75%

	Decision Tree	80	86.2	88	87.1	75%
	KNN	79	82	92.5	86.3	75%
Prostate	SVM	86.2	83.5	94.2	88.9	75%
	Decision Tree	82.5	87.1	85.3	86.2	75%
	KNN	84.3	81.2	93.8	88.1	75%

The table that has been presented provides an overview of the performance metrics that various classifiers have in terms of predicting different types of cancer. The metrics that are highlighted include accuracy, precision, recall, and F1 score, in addition to the training rates that have been utilized. The performance of prediction models for the early diagnosis of chronic diseases such as cancer can be evaluated using these measures, which serve as significant markers. It has been demonstrated that Support Vector Machine (SVM) consistently exhibits good accuracy and precision across breast, lung, and prostate cancers. This makes it a promising classifier for early detection models. The performance of decision tree classifiers is satisfactory; however, they tend to fall short of support vector machines (SVM) in terms of recall and F1 score. This is especially noticeable in the prediction of lung cancer. K-Nearest Neighbour (KNN) classifiers, on the other hand, provide results that are competitive, particularly in terms of recall, which indicates that they have the potential to be useful in identifying true positive cases. The variances that were discovered across the various types of cancer highlight the significance of adapting predictive models to particular diseases and enhancing feature selection procedures in order to improve diagnostic accuracy in early-stage detection efforts.

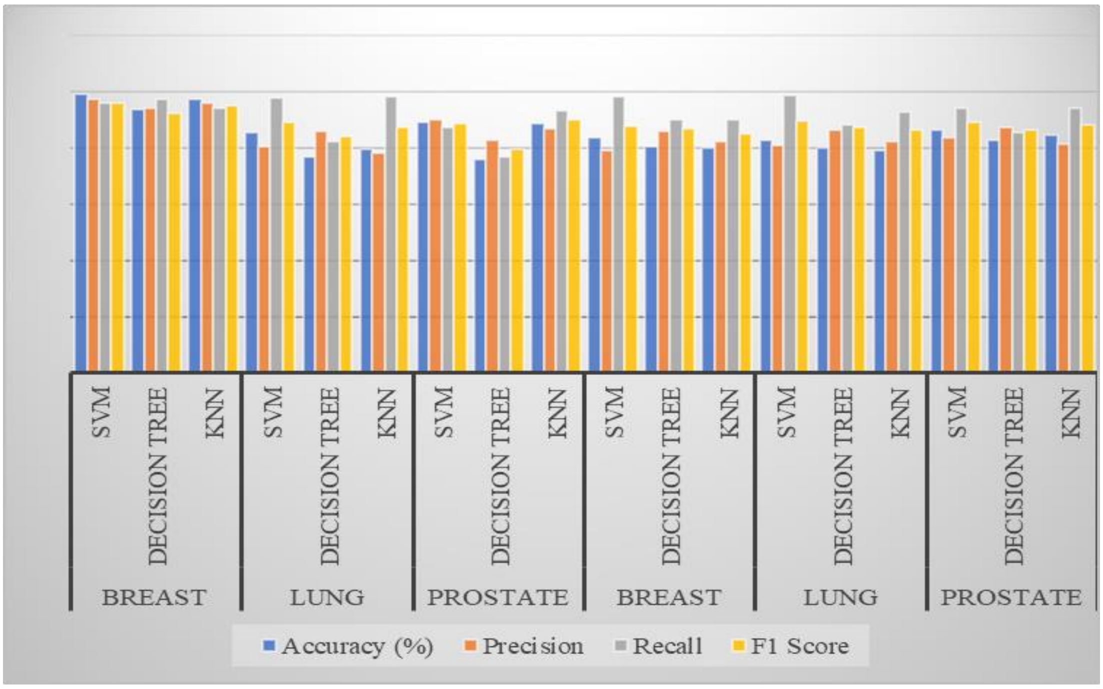

Figure:This graph illustrates the accuracy of each classifier under a variety of training rates.

5. CONCLUSION

In conclusion, the evaluation of predictive models for early diagnosis of chronic diseases such as cancer highlights the significant role that machine learning classifiers play in the healthcare industry. Upon conducting a review of performance metrics pertaining to breast, lung, and prostate cancers, it becomes evident that Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbour (KNN) classifiers exhibit a sophisticated level of effectiveness. The support vector machine (SVM) emerges as a consistently good performer, delivering great accuracy and precision, and so having promise for improving early detection efforts. In spite of this, Decision Tree and KNN classifiers also exhibit noteworthy skills, particularly with regard to particular forms of cancer or particular performance criteria. These findings highlight the significance of continuously refining and adapting predictive models, which should be informed by rigorous feature selection and customized to the specific characteristics of various chronic diseases. The ultimate goal is to improve patient outcomes by enabling earlier and more accurate diagnosis.

REFERENCES

[1] Battineni, G., Sagaro, G. G., Chinatalapudi, N., & Amenta, F. (2020). Applications of machine learning predictive models in the chronic disease diagnosis. *Journal of personalized medicine*, *10*(2), 21.

[2] Jain, D., & Singh, V. (2018). Feature selection and classification systems for chronic disease prediction: A review. *Egyptian Informatics Journal*, *19*(3), 179-189.

[3] Nusinovici, S., Tham, Y. C., Yan, M. Y. C., Ting, D. S. W., Li, J., Sabanayagam, C., ... & Cheng, C. Y. (2020). Logistic regression was as good as machine learning for predicting major chronic diseases. *Journal of clinical epidemiology*, *122*, 56-69.

[4] Kohli, P. S., & Arora, S. (2018, December). Application of machine learning in disease prediction. In *2018 4th International conference on computing communication and automation (ICCCA)* (pp. 1-4). IEEE.

[5] Pereira, S. P., Oldfield, L., Ney, A., Hart, P. A., Keane, M. G., Pandol, S. J., ... & Costello, E. (2020). Early detection of pancreatic cancer. *The lancet Gastroenterology & hepatology*, *5*(7), 698-710.

[6] Sheikhalishahi, S., Miotto, R., Dudley, J. T., Lavelli, A., Rinaldi, F., & Osmani, V. (2019). Natural language processing of clinical notes on chronic diseases: systematic review. *JMIR medical informatics*, *7*(2), e12239.

[7] Priya, R. L., & Jinny, S. V. (2021). Elderly healthcare system for chronic ailments using machine learning techniques–a review. *Iraqi Journal of Science*, 3138-3151.

[8] Landier, W., Skinner, R., Wallace, W. H., Hjorth, L., Mulder, R. L., Wong, F. L., ... & Hudson, M. M. (2018). Surveillance for late effects in childhood cancer survivors. *Journal of Clinical Oncology*, *36*(21), 2216.

[9] Bardhan, I., Chen, H., &Karahanna, E. (2020). Connecting systems, data, and people: A multidisciplinary research roadmap for chronic disease management. *MIS Quarterly*, *44*(1), 185-200.

[10] Liu, J., Liu, Y., Xiang, P., Pu, L., Xiong, H., Li, C., ... & Wang, X. (2020). Neutrophil-to-lymphocyte ratio predicts severe illness patients with 2019 novel coronavirus in the early stage. *MedRxiv*.

[11] Kumar, Y., Koul, A., Singla, R., & Ijaz, M. F. (2023). Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. *Journal of ambient intelligence and humanized computing*, *14*(7), 8459-8486.

[12] Singhi, A. D., Koay, E. J., Chari, S. T., & Maitra, A. (2019). Early detection of pancreatic cancer: opportunities and challenges. *Gastroenterology*, *156*(7), 2024-2040.

[13] Miller, D. D., & Brown, E. W. (2018). Artificial intelligence in medical practice: the question to the answer?. *The American journal of medicine*, *131*(2), 129-133.

[14] Luo, H., Zhao, Q., Wei, W., Zheng, L., Yi, S., Li, G., ... & Xu, R. H. (2020). Circulating tumor DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer. *Science translational medicine*, *12*(524), eaax7533.

[15] Cai, J., Chen, H., Lu, M., Zhang, Y., Lu, B., You, L., ... & Zhao, Y. (2021). Advances in the epidemiology of pancreatic cancer: Trends, risk factors, screening, and prognosis. *Cancer letters*, *520*, 1-11.

[16] Ijaz, M. F., Attique, M., & Son, Y. (2020). Data-driven cervical cancer prediction model with outlier detection and over-sampling methods. *Sensors*, *20*(10), 2809.

[17] Jamshidi, A., Pelletier, J. P., & Martel-Pelletier, J. (2019). Machine-learning-based patient-specific prediction models for knee osteoarthritis. *Nature Reviews Rheumatology*, *15*(1), 49-60.

[18] Crosby, D., Bhatia, S., Brindle, K. M., Coussens, L. M., Dive, C., Emberton, M., ... & Balasubramanian, S. (2022). Early detection of cancer. *Science*, *375*(6586), eaay9040.

[19] Kenner, B., Chari, S. T., Kelsen, D., Klimstra, D. S., Pandol, S. J., Rosenthal, M., ... & Wolpin, B. (2021). Artificial intelligence and early detection of pancreatic cancer: 2020 summative review. *Pancreas*, *50*(3), 251-279.

[20] Wang, W., & Wei, C. (2020). Advances in the early diagnosis of hepatocellular carcinoma. *Genes & diseases*, *7*(3), 308-319.

******