IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

A Survey On Genetic Inherited Cancer Detection In Human Using Machine Learning

[1]Sk Anamul Hoda
Research scholar,
The Department of Computer Science
The University of Burdwan,
West Bengal, India

[2]Dr. Abhoy Chand Mondal
Professor & Head,
The Department of Computer Science
The University of Burdwan,
West Bengal, India

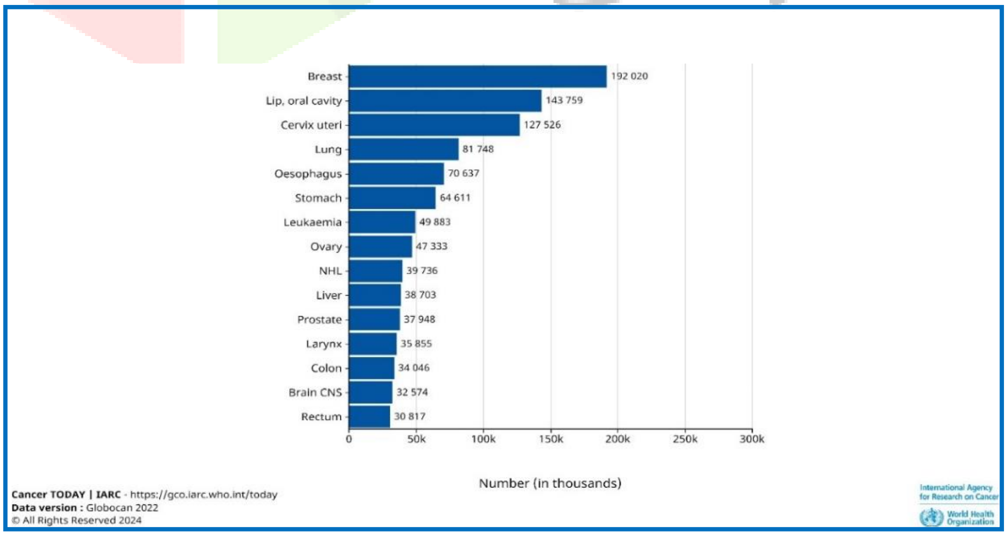
[3]Saachin Murmu
Student of M.Sc.,
The Department of Computer Science
The University of Burdwan,
West Bengal, India

Abstract: Cancer is a complicated illness that is impacted by environmental and hereditary variables. For many malignancies, inherited genetic alterations are a major contributing factor. Comprehending the hereditary foundation of cancer vulnerability is essential for timely identification, avoidance, and customized care. The goal of this research is to create a reliable method for identifying genetically inherited cancer in people. We are able to determine a person's predisposition to certain types of cancer by looking at particular genes.

Keywords: Genetic Inherited Cancer, Heredity Cancer, Machine Learning, Deep Learning

I. INTRODUCTION

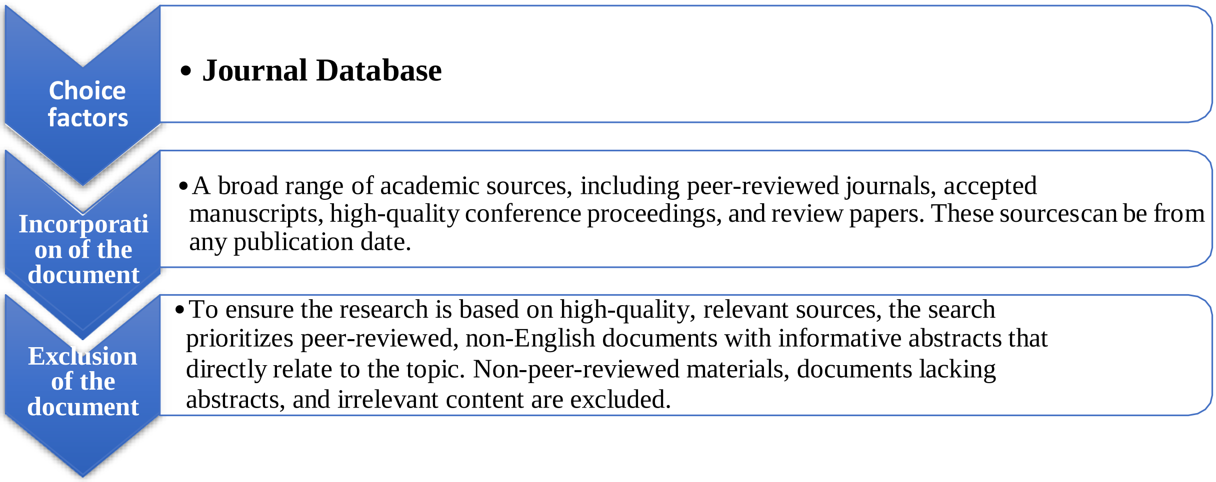
Cancer among the worst disease is becoming more common, thus early detection and treatment are essential. When a person receives particular genetic variations, often referred to as mutations, from their parents, their risk of acquiring particular cancer kinds is greatly increased. This is referred to as genetically inherited cancer. Hereditary cancer is another term for genetically inherited cancer. Figure 1 depicts the number of cancer cases in India in 2022.



Cancer TODAY | IARC - https://gco.iarc.who.int/today
Data version : Globocan 2022
© All Rights Reserved 2024

Figure 1: Cancer cases in 2022 in India

The mutation responsible for inherited cancer occurs in the sperm or egg cells (germline cells) at conception. This means the mutation is present in every cell of the developing embryo and the resulting adult. These mutations often disrupt genes involved in cell growth and repair.



- **Journal Database**

- A broad range of academic sources, including peer-reviewed journals, accepted manuscripts, high-quality conference proceedings, and review papers. These sourcescan be from any publication date.

- To ensure the research is based on high-quality, relevant sources, the search prioritizes peer-reviewed, non-English documents with informative abstracts that directly relate to the topic. Non-peer-reviewed materials, documents lacking abstracts, and irrelevant content are excluded.

Figure 2: Methods involved for conducting a review

While inheriting the mutation doesn't guarantee cancer, it significantly increases a person's risk compared to the general population. Well-known examples of inherited cancer genes include BRCA1, BRCA2, ATM, CHEK2 etc. which raise the risk of prostate, pancreatic, breast andovarian cancer. It's important to remember that inheriting a cancer susceptibility gene isn't thesole factor. Other lifestyle and environmental influences can also play a role. However, for those with a family history of cancer or those who were diagnosed at an early age, genetic testing can offer useful information. Many lives and millions of rupees could be saved by early cancer diagnosis. Early cancer detection and prognosis are critical to cancer research and treatment. Treatment for cancer will be very effective if it is discovered early. By closely observing hundreds of cancer genes using gene expression data, cutting-edge technology will significantly advance the early diagnosis of hereditary malignancies.

II. METHODOLOGY

The goal of the research subject review is to investigate relevant studies and techniques that are used to detect genetically inherited cancers.

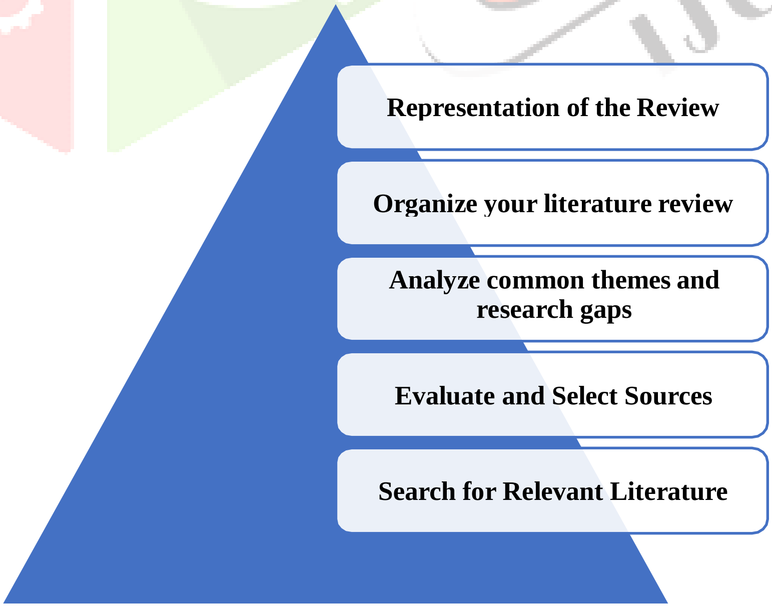


Figure 3: Selection criteria for journal databases

Hereditary malignancies are defined as cancers, brought on by an inherited mutation. Sporadic cancers are the remainder of the malignancies that are not brought on by an inherited mutation. Although most cancers are sporadic, an inherited mutation is thought to be the cause of 10–20% of malignancies worldwide. Understanding hereditary cancers still heavily depends on a person's family history of the disease. Although genetic alterations or mutations within the body are the root cause of all malignancies, a certain proportion of these abnormalities are inherited. The remaining ones happen on their own, and it appears that the environment has a big influence. It's crucial to understand that possessing a gene variant connected to cancer does not ensure that a person will get cancer, particularly in families where the variant is shared. But it increases the likelihood of the illness striking the person. The table below lists some common cancer types along with the corresponding hereditary mutations.

Table 1: Common cancer types and associated inherited gene mutations

Forms of Hereditary Cancer	**Genomic mutations list**
Breast	ATM, BARD1, BRCA1, BRCA2, CDH1, CHEK2, PALB2, PTEN, RAD51C, RAD51D, STK11, TP53
Colorectal	APC, CHEK2, EPCAM, MLH1, MSH2, MSH6, MUTYH, PMS2, PTEN, TP53
Endometrial	BRCA1, EPCAM, MLH1, MSH2, MSH6, PMS2, PTEN, STK11, TP53
Fallopian tube, Ovarian	BRCA1, BRCA2, BRIP1, EPCAM, MLH1, MSH2, MSH6, PALB2, PMS2, RAD51C, RAD51D
Gastric	APC, CDH1, EPCAM, MLH1, MSH2, MSH6, MUTYH, STK11
Melanoma	BRCA2, CDK4, CDKN2A, PTEN
Pancreatic	ATM, BRCA1, BRCA2, CDKN2A, MLH1, PALB2, STK11, TP53
Prostate	BRCA1, BRCA2, CHEK2, EPCAM, HOXB13, MLH1, MSH2, MSH6, PMS2, TP53

Cancers that are inherited are those brought on by a genetic mutation that existed in the sperm or egg cell during conception. These malignancies comprise a portion of less frequent cancers such as ovarian and pancreatic cancer, as well as prevalent cancers such as breast, colon, and prostate cancer. It's crucial to remember that the existence of the mutation does not guarantee the development of cancer. Listed below are a few well-known genetic mutations that can be inherited by offspring:

- One gene associated with cancer is known as ATM. A family history of cancer may result from inherited mutations in ATM, which also raise the risk of prostate, pancreatic, female breast, and perhaps other cancers.
- One gene that has been connected to cancer is called BRCA1. A family history of cancer can result from inherited mutations in BRCA1, which also raise the chance of ovarian, pancreatic, prostate, and possibly other cancers in men and women.
- One gene associated with cancer is called BRCA2. Male and female breast cancer, ovarian, pancreatic, prostate, melanoma, and probably other cancers are all at increased risk due to inherited mutations in BRCA2, which can also cause cancer to run in families.
- One gene associated with cancer is called CHEK2. Breast cancer (in women and perhaps in males), colorectal cancer, and possibly prostate cancer are all made more likely by inherited mutations in CHEK2. Cancer may run in families when CHEK2 mutations occur.
- One gene connected to cancer is called PALB2. Mutations in PALB2 that are inherited raise the chance of pancreatic, ovarian, and female breast cancers in addition to possibly other cancers. These mutations can also make cancer run in families.
- One gene associated with prostate cancer is called HOXB13. Prostate cancer is more likely to occur in families when inherited mutations in HOXB13 are present.

Artificial intelligence (AI) has a subset called machine learning (ML) that gives machines the capacity to learn from data without the need for explicit programming. Vast amounts of data are used to train machine learning systems. By examining patterns and trends in the data, the algorithms can function better on data that will be observed in the future. Predictiveness is the fundamental function of machine learning. ML algorithms are autonomous, as opposed to traditional programming, where each step is explicitly stated. This enables them to address difficult issues where it would be impractical or even impossible to define every rule. Within the ML field, the technique known as "supervised learning" was developed using a labelled dataset, in which each sample was linked to a certain label or result. To enable the algorithm to generate predictions on fresh, unobserved data, the objective is to develop a mapping from input features to output labels. Datasets used in supervised machine learning are usually divided into training partitions and testing partitions. The ML algorithm receives its training from the training set.

The trained algorithm's accuracy is assessed using the testing data.

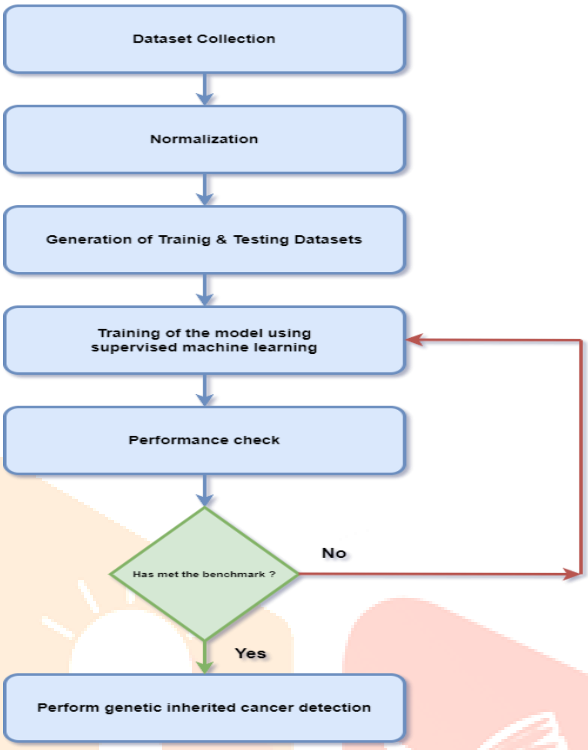


Fig 4: Diagram of the proposed AI Model

Figure 4 depicts the suggested machine learning-based genetically inherited cancer detection approach. The first step is gathering and organizing information into a database on inherited genes and the tumors they cause. The second stage will involve data preprocessing to eliminate duplicate data using various queries and normalize data using various normalization techniques. The data will be split into training and testing categories in the third phase and kept in different databases. Using supervised machine learning techniques and training data, the suggested AI model will be trained in the fourth step. If the trained model's performance meets the benchmark, the AI model will be evaluated in the fifth phase; if not, it will need to be retrained. In the last stage, the suggested AI model will carry out inherited cancer diagnosis by importing the testing data from the database.

III. LITERATURE REVIEW

A form of genetic testing known as genetic inherited cancer detection scans for gene mutations that may raise a person's chance of getting a particular cancer. Children may inherit these mutations from their parents. Algorithms have become more and more significant in a variety of disciplines, and the field of machine learning has grown significantly in recent years. This article examines the application of ML methods for diagnosing cancer that is genetically inherited. To start, we give a quick summary of the research issue and its importance.

The authors of [1] detailed the association between genes 1 (BRCA1) and 2 (BRCA2) and the incidence of germline mutations that increase the risk of developing breast and ovarian cancer. Significant percentages of stomach, pancreatic, and prostate cancers, as well as advanced disease vulnerability, were caused by mutations in the BRCA1/BRCA2 genes.

According to [2], data mining has the potential to be a useful tool for diagnosing, prognosticating, and treating breast cancer cases by finding commonalities (patterns) in the cases. ANN, SVM and Decision Tree (DT) were the three methods they covered. By using information from the Iranian Center for Breast Cancer, the researchers were able to compare various methods. The most accurate method, they discovered, was the support vector machine method.

The authors of [3] noted that developing precise and computationally effective classifiers for medical applications is a significant challenge in the fields of data mining and machine learning. Using Wisconsin Breast Cancer (original) datasets for both diagnosis and analysis, they examined the behaviour of data mining methods SVM, NB, K-NN, and Decision Tree (C4.5). They discovered that SVM performed the best in terms of accuracy and low error rate.

In [4], the writers contrasted methods of machine learning for predicting prostate cancer. It talks about how difficult it is to diagnose prostate cancer and how improved diagnostic techniques are required. In situations where there are no set rules, machine learning techniques are believed to be helpful. On a dataset comprising one hundred patients, the authors examine different supervised machine-learning methods. The best accuracy can be obtained with a MLP classifier, according to the results.

In **[5]** author discussed about prediction prognosis of pancreatic cancer patients after surgery. The article discusses the role of mucin genes in cancer. Mucin expression levels and DNA methylation status can be used to predict prognosis. A machine learning model was developed to analyze these factors. The model was able to distinguish between patients with high and lowrisk of death. This suggests that the analysis of mucin genes could be a useful tool for the prognosis of pancreatic cancer.

In **[6],** using gene expression data, the author suggests a deep learning technique to overcome the problem of high dimensionality in cancer categorization. The method utilizes feature selection techniques and machine learning algorithms on microarray data. Convolutional neural networks were applied to lung, kidney and brain cancer datasets.

In **[7]** authors introduced a new machine learning model for classifying leukaemia. They discussed two types of leukaemia: acute lymphocytic (ALL) and acute myelocytic (AML). The model uses deep neural networks (DNN) to classify gene expressions. The authors claimed thattheir model is more accurate than other models.

In **[8]** authors explained deep learning for cancer type classification and driver gene identification. They discussed using raw cancer DNA sequencing data to classify cancer typesand identify relevant genes. The authors propose a method called DeepCues, which is a convolutional neural network model. DeepCues can also be used to identify cancer-relevant genes. In **[9]** the authors explained how rule-based ML can be used to categorize, collection of data on cancer gene expression and prioritize putative disease genes. They discussed several machine learning techniques to achieve high accuracy in classifying cancer samples, while also providinginterpretable results.

In **[10]** authors researched about evaluating machine learning methods for identifying cancerdriver genes. They discussed about the importance of recognizing these genes for accurate cancer analysis.

In **[11]** authors explored different machine learning methods for cancer classification usinggene expression data. They discussed different types of gene expression data, including microarrays and RNA-Seq. Machine learning algorithms are used to analyze this data to classifydifferent types of cancer.

In **[12]** authors developed a deep learning approach for cancer subtype classification. It discusses using high-dimensional gene expression data to classify cancer subtypes. Currentmethods struggle with this task because the data is sparse and high-dimensional.

In **[13].** gene expression data classification was explained by the authors using SVM combined with mutual information-driven gene selection. It discusses microarray data analysis and its use in classifying diseases. The authors propose a method for gene classification that utilizes SVM and mutual information (MI) for gene selection.

The use of machine learning techniques for cancer driver gene prediction was covered by authors in [14]. Itdiscusses the challenges of distinguishing between driver and passenger mutations. The authors reviewed different machine learning algorithms used for cancer driver prediction. Italso summarizes the types of data used to train these models. Data from several sources can be integrated by ML models.

In **[15]** authors reviewed about utilizing ML for cancer prediction and prognostic analysis. It discusses machine learning methods used to analyze cancer data. ML is a type of AI that can infer from data. This article focuses on itsuse in cancer prognosis. Prognosis refers to the likely outcome of a disease. Machine learningcan be used to predict a patient's likelihood of developing cancer, the chance of the cancer recurring, and how long a patient might live. Doctors can use this information to make treatment decisions.

IV. ANALYSIS OF LITERATURE REVIEWED

Our analysis of Tables 2 and 3 revealed two key trends in genetically inherited cancer detection.First, a bar chart based on data from Table 3 highlights the various application areas where thisdetection is crucial. Second, a column chart derived from Table 4 showcases the prevalence of different machine learning techniques used in this field. Notably, the column chart indicates that supervised machine learning methods are favoured for genetic inherited cancer detection.

Table 2: Highlights and recommendations for future directions from the previous works

Related Paper	**Prominent feature**	**methods employed**	**Use case/disease**	**Noted Challenges**	**Recognized shortcomings**	**Upcoming research path**
[1]	Explores connections between BRCA1/2 mutations and cancers of the prostate, pancreas and stomach	Gene mutation	Prostate, pancreatic and stomach cancer detection	Conflicting findings regarding the association of BRCA1/2 gene mutation with these cancers	Findings are inconclusive, which means more research is needed to determine how BRCA1/2 mutations influence these cancers	More research is needed to determine how BRCA1/2 mutations influence these cancers
[2]	Using three machine learning techniques for	Machine learnin	Prediction of breast cancer	Each ML algorithm has its limitations and strengths specific to	There were many cases lost in the follow-up and there were records	The performance of the models can

	predicting breast cancer recurrence	g	recurrence	the type of application	with missing values that were omitted unfortunately	be enhanced
[3]	Access the correctness in classifying data concerning the efficiency and effectiveness of each algorithm in terms of accuracy, sensitivity and specificity	Machine learning	Breast cancer risk prediction and diagnosis	Constructing accurate and computationally efficient classifiers for medical applications	Simulation errors are also considered in this study	Machine learning techniques can be applied to detect genetically inherited cancers
[4]	Comparing machine learning approaches to predict prostate cancer	ML	Prediction of prostate cancer	The low predictive performance of the diagnostic techniques now in use and the absence of defined criteria for assessing the symptoms of prostate cancer	absence of precise guidelines for assessing the symptoms of prostate cancer. This complicates the disease's diagnosis.	Use a larger dataset or other machine learning techniques

Related Paper	**Prominent feature**	**methods employed**	**Use case/disease**	**Noted Challenges**	**Recognized shortcomings**	**Upcoming research path**
[5]	Role of		Forecasting	Even with endoscopic ultrasound-guided fine-needle	This study did not include other biomarkers that may improve	Use this method for other types

	mucin genes in pancreatic cancer	Machine learning	the prognosis of pancreatic cancer patients	aspiration, differentiate between pancreatic cancer and other pancreatic disorders such chronic pancreatitis can be challenging.	the accuracy of the model. The authors acknowledge that other markers are being identified and the model could be improved by including them	of cancer or to investigate other genes that might be involved in prognosis
[6]	Deep learning technique for classifying cancer profiles using gene expression data	Deep learning	Detection of lung, kidney, and brain cancers	The high dimensionality of data in machine learning. This means that there are too many variables to analyze	Data sparsity. This means that there is a lot of missing data	Exploration of other unsupervised feature selection techniques or investigate supervised feature selection techniques
[7]	Classifying leukaemia from gene expression	Machine Learning	Diagnosis of leukaemia	Selecting suitable features from the original data. Dealing with the high dimensionality of the data	The dataset is relatively small to confirm its generalizability	Improve the accuracy of the model by optimizing the network structure
[8]	Cancer type classification and driver gene identification	Deep learning	Prediction of brain, breast, colorectal, kidney, lung, prostate, and uterus cancers	The high dimensionality of features when studying somatic mutations, germline variants, insertions	Future development is needed to better evaluate the model	Accuracy can be improved by using a larger dataset or a more complex model

				and deletions together		architecture

Related Paper	**Prominent feature**	**methods employed**	**Use case/disease**	**Noted Challenges**	**Recognized shortcomings**	**Upcoming research path**
[9]	Prioritization of candidate disease genes and classifying samples of cancer gene expression data	Machine learning	Detection of breast, prostate, and lymphoma cancers	Small sample sizes, a high percentage of uninformative genes, high noise levels, several outliers, and systematic bias were the problems.	CFS method may not be the most appropriate choice in all situations, and the optimality of any given method depends on the specific dataset	Develop new machine learning techniques that are even more accurate and interpretable
[10]	Evaluation of machine learning algorithms for identification of cancer driver genes	Machine learning	Identification of cancer driver genes	Efficiently combining and optimizing numerous methodologies to identify cancer driver genes on large datasets	Mainly relied on frequency based criteria to prioritize transformations	More dependable methods are needed for identifying cancer driver genes
[11]	Classifying cancer based on gene expression data	Deep learning	Detection of breast, lung, ovarian, colon, prostate cancers and leukaemia, myeloma	Generalizability, robustness, and explainability of the results	Small sample size with high dimensionality	Develop methods to handle the challenges of small datasets and high dimensionality

[12]	Cancer subtype classification	Deep learning	Identification of breast and bladder cancer	Cancer samples are scarce and the high-dimensional features of their gene expression data are too sparse	Only used gene expression data, and does not take into account other genomic data sources	Future research will look into incorporating multiomics data

Related Paper	Prominent feature	methods employed	Use case/disease	Noted Challenges	Recognized shortcomings	Upcoming research path
[13]	Gene expression data classification and mutual information-based gene selection	Machine learning	Detection of colon cancer and lymphoma	Gene expression data classification, which includes high dimensionality and noisy or irrelevant genes	Selection of informative genes	Explore the use of ensemble methods to combine the strengths of multiple gene selection and classification algorithms
[14]	Reviews of several ML techniques for the prediction of genes that cause cancer	Machine learning	Prediction of cancer driver genes	Methodological differences among different methods can cause inconsistencies in the outputs	It is possible to increase the precision of ML techniques for identifying cancer driver genes.	Develop new methods that can integrate more data types
[15]	Leveraging ML for cancer prediction and prognosis	Machine learning	Cancer prediction and prognosis	Large amounts of data are needed to train machine learning models. Patient data is often	Studies need to use more training data	Improve the accuracy of cancer prediction and prognosis

				privacy sensitive and can be difficult to collect		

Table 3: a synopsis of the many reviews application areas

Referred paper	**Use case/cancer**	**Total**
[2] [3] [8] [9] [11] [15]	Breast cancer	6
[1] [4] [8] [9] [11] [15]	Prostate cancer	6
[1] [5]	Pancreatic cancer	2
[1]	Stomach cancer	1
[11]	Ovarian cancer	1
[6] [8] [11] [15]	Lung cancer	4
[6] [8] [11]	Kidney cancer	3
[6] [8]	Brain cancer	2
[11] [15]	Liver cancer	2
[11] [13]	Colon cancer	2
[7] [11]	Leukaemia	2
[8] [11]	Uterus cancer	2
[8] [15]	Colorectal cancer	2
[9] [11] [13] [15]	Lymphoma	4
[15]	Skin cancer	1
[15]	Throat cancer	1
[12] [15]	Bladder cancer	2

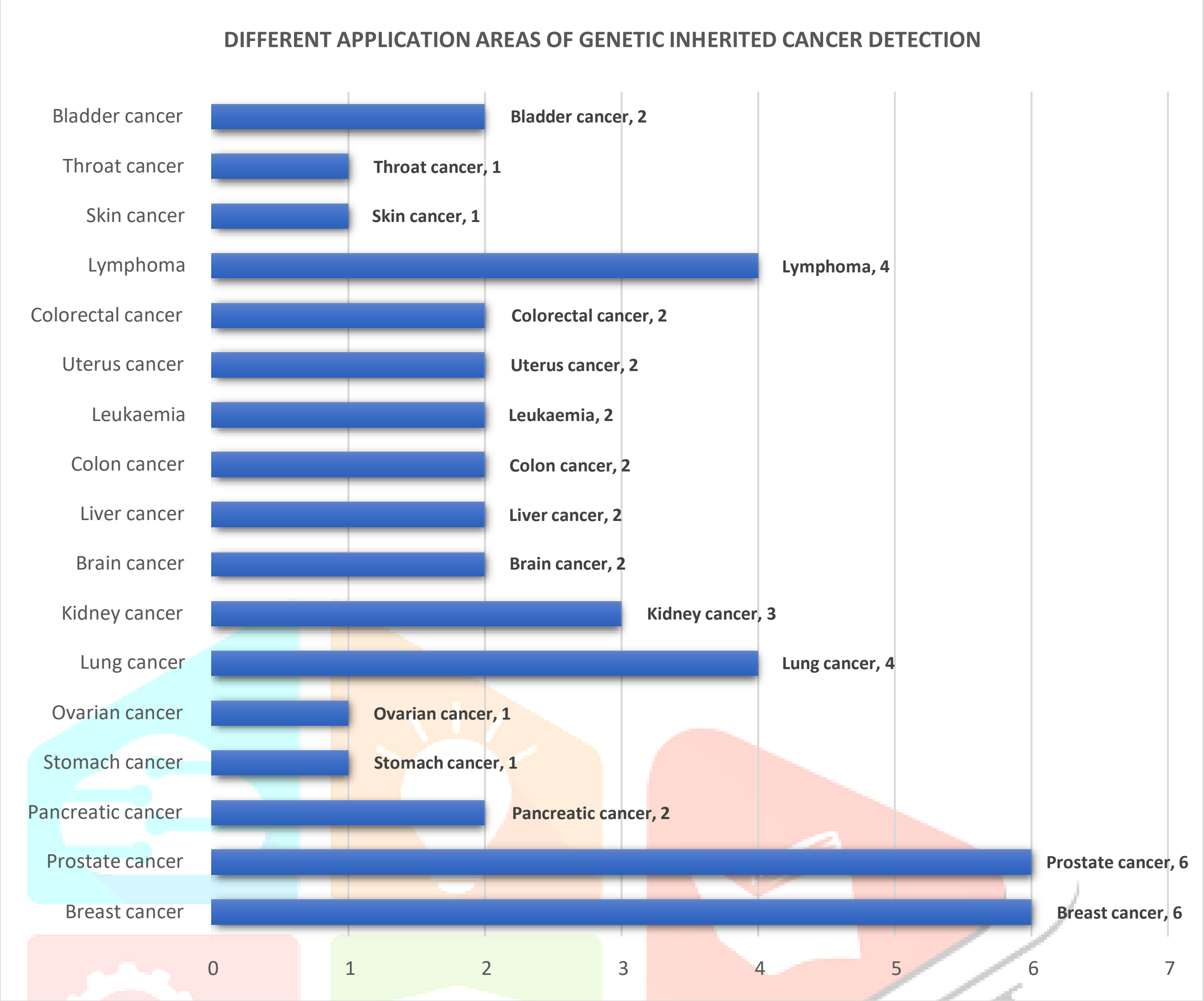


Fig 5: Different application areas of Genetic Inherited Cancer Detection

Table 4: Machine learning techniques used in genetic inherited cancer detection

Referred paper	**Machine learning techniques/methods**	**Count**
[2] [3] [4] [5] [7] [9] [10] [11] [12] [13] [14] [15]	Support Vector Machine (SVM)	12
[2] [5] [10] [13] [14] [15]	Artificial Neural Network (ANN)	6
[6] [8] [11] [12] [14]	Convolutional Neural Networks (CNN)	5
[11] [14]	Graph Neural Networks (GNN)	2
[11]	Transformer Neural Networks (TNN)	1
[11]	Recurrent Neural Networks (RNN)	1
[4] [7]	Deep Neural Network (DNN)	2
[2] [3] [11] [12] [15]	Decision Tree (DT)	5
[3] [4] [7] [14] [15]	Naive Bayes (NB)	5
[3] [4] [7] [11] [13] [15]	K Nearest Neighbor (K-NN)	6
[4] [9] [10] [11] [12] [14]	Random Forest (RF)	6
[4]	Linear Regression	1

[4] [13] [14]	Logistic Regression	3
[4] [12]	Multi Layer Perceptron (MLP)	2
[13] [15]	Fuzzy logic	2

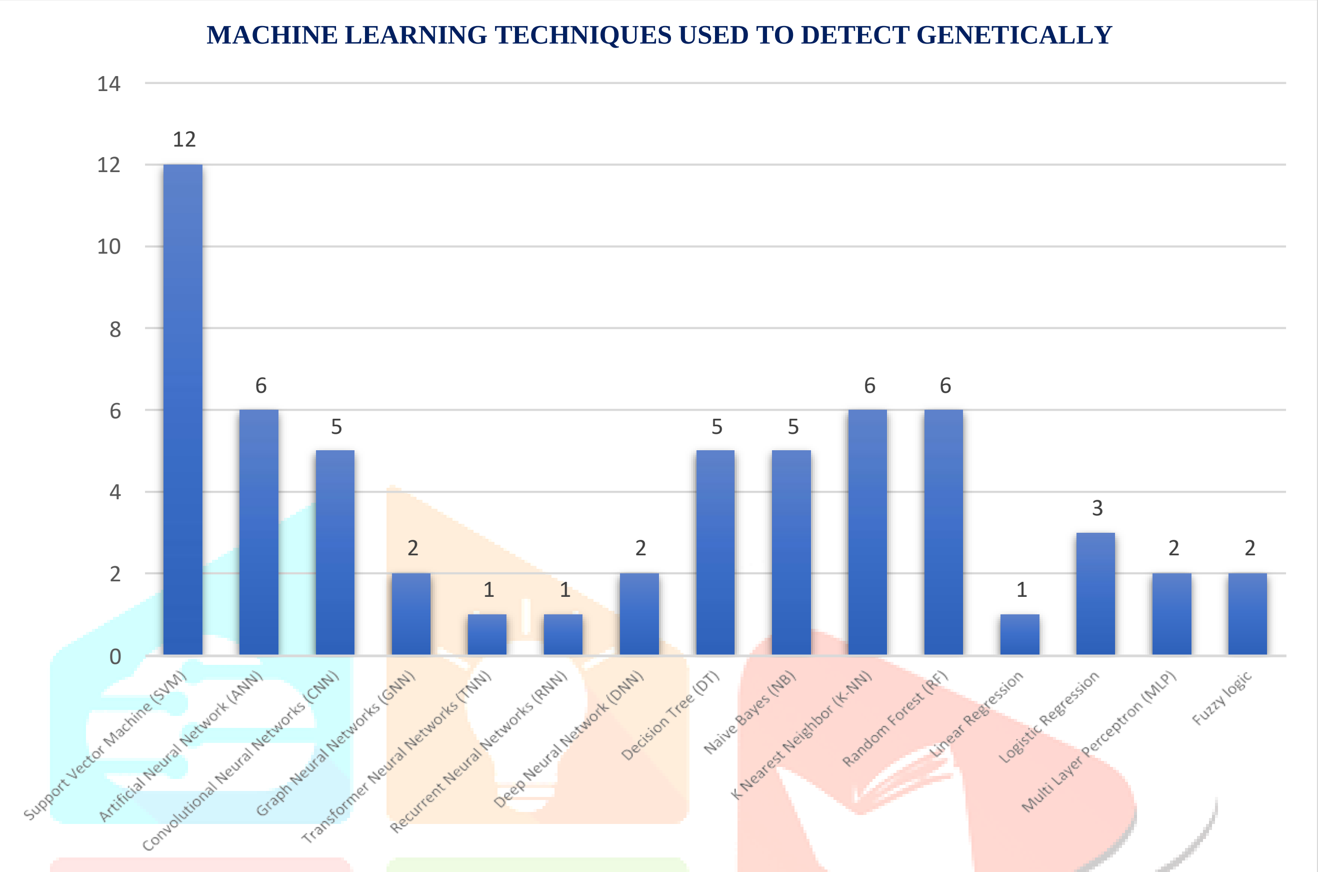


Figure 6: Machine learning techniques used to detect genetically inherited cancers

V. CONCLUSION AND FUTURE DIRECTIONS

This research investigates the role of gene mutations in identifying inherited cancers through machine learning. We reviewed existing algorithms to emphasize the importance of choosing the right method for different cancers. Since inherited cancer detection relies on family history and patient data, ethical considerations are paramount. Patient data privacy is a concern, so we must identify potential privacy breaches and misuse of information. Obtaining documented ethical approval is essential before starting the study.

Machine learning and doctors bring unique strengths to medical diagnosis. While doctors leverage their knowledge and experience to interpret new information, machine learning excels at analyzing vast amounts of historical data to identify patterns. Machine

learning can't replace doctors, but it serves as a powerful tool to accelerate diagnoses, shorten therapy duration, reduce the workload on medical staff, and optimize resource allocation.

The primary aim of this review was to underscore the significance of integrating machine learning algorithms for the early detection of genetically inherited cancers.

VI. ACKNOWLEDGMENT

We gratefully acknowledge the support and technical assistance provided by the Department of Computer Science at the University of Burdwan, West Bengal, India.

REFERENCES

[1] Cavanagh, H. & Rogers, K. M. (2015) The role of BRCA1 and BRCA2 mutations in prostate, pancreatic and stomach cancers. Hered Cancer Clin. Pract. 13, 16. https://doi.org/10.1186/s13053-015-0038-x

[2] Ahmad LG, Eshlaghy AT, Poorebrahimi A, et al. (2013) Using three machine learning techniques for predicting breast cancer recurrence. Journal of Health and Medical Informatics 4(124): 3–5 https://doi.org/10.4172/2157-7420.1000124

[3] Asri H, Mousannif H, Al Moatassime H, et al. (2016) Using machine learning algorithms for breast cancer risk prediction and diagnosis. Procedia Computer Science 83: 1064–1069. https://doi.org/10.1016/j.procs.2016.04.224

[4] Ebru Erdem, Ferhat Bozkurt (2021) A Comparison of Various Supervised Machine Learning Techniques for Prostate Cancer Prediction https://doi.org/10.31590/ejosat.802810

[5] Michiyo Higashi (2020) Predicted Prognosis of Patients with Pancreatic Cancer by Machine Learning https://doi.org/10.1158/1078-0432.CCR-19-1247

[6] Hatim Z Almarzouki (2022) Deep-Learning-Based Cancer Profiles Classification Using Gene Expression Data Profile https://doi.org/10.1155/2022/4715998

[7] Pradeep Kumar Mallick, Saumendra Kumar Mohapatra, Gyoo-Soo Chae, Mihir Narayan Mohanty (2020) Convergent learning–based model for leukemia classification from gene expression https://doi.org/10.1007/s00779-020-01467-3

[8] Zexian Zeng, Chengsheng Mao, Andy Vo, Xiaoyu Li, Janna Ore Nugent, Seema A. Khan, Susan E. Clare7, Yuan Luo (2020) Deep learning for cancer type classification and driver gene identification https://doi.org/10.1186/s12859-021-04400-4

[9] nrico Glaab, Jaume Bacardit, Jonathan M. Garibaldi, Natalio Krasnogo (2012) Using Rule- Based Machine Learning for Candidate Disease Gene Prioritization and Sample Classification of Cancer Gene Expression Data https://doi.org/10.1371/journal.pone.0039932

[10] Sharaf J. Malebary, Yaser Daanial Khan (2021) Evaluating machine learning methodologies for identification of cancer driver genes https://doi.org/10.1038/s41598-021-91656-8

[11] Fadi Alharbi, Aleksandar Vakanski (2023) Machine Learning Methods for CancerClassification Using Gene Expression Data: A review https://doi.org/10.3390/bioengineering10020173

[12] Jiquan Shen, Jiawei Shi, Junwei Luo, Haixia Zhai, Xiaoyan Liu, Zhengjiang Wu, ChaokunYan, Huimin Luo (2022) Deep learning approach for cancer subtype classification using high-dimensional gene expression data https://doi.org/10.1186/s12859-022-04980-9

[13] Devi Arockia Vanitha C, Devaraj D, Venkatesulu M (2015) Gene Expression Data Classification using Support Vector Machine and Mutual Information-based Gene Selection https://doi.org/10.1016/j.procs.2015.03.178

[14] Renan Andrades, Mariana Recamonde-Mendoza (2022) Machine learning methods forprediction of cancer driver genes: a survey paper https://doi.org/10.1093/bib/bbac062

[15] Joseph A. Cruz, David S. Wishart (2006) Applications of Machine Learning in CancerPrediction and Prognosis https://doi.org/10.1177/117693510600200030