Streamlining Cancer Diagnosis and Prognosis System using Hybrid CNN-NPR: Deep Learning Approaches

A. Ganesh[1], SivaKumar Depuru[2], Basi Reddy A.[3], G. Sujatha[4]

Abstract: In the context of medical science, advancements in technical infrastructure related to computer and life sciences have enabled the utilization of computational methods for medical diagnosis. As the number of cancer cases continues to rise rapidly, the existing diagnostic system is becoming out dated, necessitating the development of modern, productive, and optimized strategies. Accurately predicting the type of cancer is crucial for the diagnosis and treatment of the disease. Knowledge of cancer genes can significantly assist in comprehending, diagnosing, and identifying the different types of cancer. In this study paper, the identification and prediction of cancer type are achieved through the utilization of hybrid CNN-NPR several researchers have proposed different Convolutional Neural Network (CNN) models to date. Every model focused on a specific group of parameters that were utilized to imitate the gene pattern.

Keywords: CNN, Deep Learning, CNN-NPR, Cancer diagnosis, Prognosis system

1. Introduction

The medical and health sector is currently a major area of research in the 21st century, with a significant focus on enhancing life span [1]. In today's challenging healthcare market, healthcare facilities must prioritize quality improvement activities to grow and stay competitive [2,3]. To achieve this goal, healthcare systems need to be capable of recognizing the actions that will yield the greatest positive effect on their financial performance. Programs aimed at improving quality, whether focusing on clinical, financial, or operational aspects, can significantly impact the total expense of healthcare, patient results, care consistency, support for clinical decision-making, duration of hospitalization, and numerous other factors. [4,5]. Hospitals across the nation face several critical issues, including practice variation, preventable medical mistakes, prolonged hospital stays, and financial deterioration [6,7]. The progress of science, along with progress in telecommunications infrastructure and information technology, has brought about a greater emphasis on precise diagnostic and predictive methods. This holds the potential to offer the assurance of a promising and healthy life ahead [8]. Technological advancements have the potential to greatly enhance the overall quality of life; in the initial phases of creating proactive detection methods, numerous computational techniques and software were utilized. Scientific progress, along with the evolution of information

exchange systems and digital innovations, has led to an increased emphasis on diagnostic and predictive methodologies to ensure a promising and healthy future. As technologies continue to advance and radiological techniques evolve, diagnostic imaging techniques will progressively enhance their precision and reliability. Now, in addition to the healthcare sector and concerns, scholars are also prioritizing the diagnostic process to improve outcomes effectively. Major tech giants such as Microsoft and Google have shifted their attention towards developing advanced computational methods to address the crucial issue of early-stage cancer diagnosis and prediction [9]. Microsoft's substantial allocation of resources towards cloud computing is a logical step for a field that demands ample computational capabilities to address complex problems. One approach center around the concept that diseases such as cancer can be viewed as information processing systems. As per Jeanette M. Wing, the corporate vice president responsible for Microsoft's basic research labs, the company's approach to curing cancer is centered around two core methods, reflecting their commitment to sustainability. Google claims that their system has the potential to detect a higher number of false negatives compared to previous studies, specifically in seemingly normal photos that actually exhibit indications.

The complexity of cancer, acknowledged universally, encompasses various subtypes of the disease. The importance of early cancer research lies in its impact on improving patient medical treatment, primarily focusing on prompt cancer identification and treatment [10,11]. Various research collectives have delved into applying machine learning (ML) and deep learning techniques within the realms of chemical biology. Their objective is to distinguish cancer patients into high or low-risk classifications. [12,13].

[1] *Dept.of. Computer Science and Engineering, Sri Venkateswara College of Engineering, Tirupathi, India*
[2] *School of Computing, Mohan Babu University, Tirupathi, India. sivakumar.d@vidyanikethan.edu*
[3] *School of Computing, Mohan Babu University, Tirupathi, India*
[4] *Dept.of. Electronics and Communications Engineering, Sri Venkateswara College of Engineering, Tirupathi, India*

Within the domains of medical science and computational biology research, a critical effort involves the thorough investigation by various research teams into employing machine learning (ML) techniques for the categorization of cancer patients, ensuring effective risk assessment. Researchers have utilized such methodologies with the aim of simulating the progression and treatment of cancerous diseases. The importance of these studies is underscored by the ability of ML algorithms to identify crucial features within complex datasets. In recent years, the utilization of machine learning has gained prominence in the field of cancer research, specifically for the purposes of forcasting and prediction. This alternative approach is particularly captivating as it aligns with the emerging trend towards customized and anticipatory advancements in medical care. Consequently, cancer development and treatment have been influenced by these strategies, leading to the adoption of similar approaches. The ability of machine learning algorithms to identify significant characteristics in extensive datasets is of utmost importance. Numerous approaches are frequently employed to construct prediction techniques that anticipate the emergence of a potential cancer cure [14,15].

Statistically, among every six fatalities, one is attributed to cancer, making it one of the most lethal and second leading causes of death [16]. An essential contributing factor to this phenomenon is the heightened complexity and systemic nature of predictive models for prostate and breast cancer, which often surpass those developed for lung cancer. Establishing a dependable system for predicting cancer in its early stages holds significant importance. [2,5]. Scientists are actively engaged in the prediction of cancer by discerning unique markers generated by specific genes associated with each and every type of cancer. This enables the construction of a learning mechanism based on patterns for accurate identification. By leveraging innovative approaches for early detection, advancements can play a crucial role in enhancing the quality of life and promoting better health outcomes. Tumour biomarkers for cancer diagnostics have primarily faced limitations in terms of inadequate sensitivity for cancer detection and precision in distinguishing cancer. Conversely, most tumor markers are subject to limitations in terms of their ability to specifically identify organs. The utilization of tumor markers in early detection, assessment, or monitoring of cancer relapse can provide doctors with insights into the extent of cancer progression within the body. During the initial stages of developing pre-detection methods, various computational approaches and tools were employed [17,18]. The initial successful method involves employing a simulated neural network that emulates the functionalities of neurons in the human cerebral system. [19]. The fundamental building blocks of this system are specialized cells called "neurons," which facilitate the connection between input and output by means of multiple interconnected layers [20]. Rapid and accurate detection is crucial when it comes to one of the most severe illnesses, brain tumors. Upon acquiring the MRI image data, machine learning algorithms are employed at various stages within automated detection systems to assist diagnosis. This approach employs a model based on artificial neural networks to analyzes the interplay between genetic factors, nutritional aspects, and demographic variables in order to predict an individual's likelihood of developing breast cancer. This approach has the potential to offer valuable understanding into an individual's susceptibility to the condition even prior to its manifestation. In the realm of cancer detection, a neural network undergoes two distinct phases: adaptation and verification. Initiated with a dataset, the network undergoes its initial training, ensuring comprehensive learning. Afterward, the network undergoes validation to evaluate classifications on a new dataset, once the weights in the connections between neurons have been established.Exploration involved employing a machine learning model for the task of automatic categorization—the ANN. Within the realm of medical image analysis, artificial neural networks (ANNs) find application in categorizing and identifying patterns across varied datasets. They have demonstrated potential as a technique for accurately categorizing breast cancer. ANN is utilized for early detection of breast cancer in various imaging modalities, including mammography, ultrasound, MRI, and infrared imaging. The artificial neural network (ANN) consistently examines the intermediate layer alongside the intended result in order to minimize the discrepancy in the perception process. Thus, it strives to diminish the error through ongoing comparisons. As a result of the limited precision, the researcher redirected their attention towards more precise and sophisticated methodologies such as advanced neural networks. A component of artificial intelligence known as deep learning involves neural networks with numerous layers, amplifying its ability for sophisticated learning and producing more accurate results through the utilization of state-of-the-art methods like big data analysis. The process commences with the initial training phase, wherein the determination of model parameters is accomplished by utilizing datasets. This step involves the training of the model using diverse data sources to derive optimal parameter values [8,14]. With proper instruction and extensive training, the forecasting is conducted with precision as it encompasses numerous interconnected layers and leverages vast amounts of data. The effectiveness of a system depends on the amount of data used for network training.[21]. Figure 1 of techniques utilized for predicting cancer.

Fig 1: Prognosis of Cancer using ANN

A neural network operates resembling an adaptable mechanism, wherein it modifies its configuration during the learning process. Neural networks offer a convenient means of articulating both straightforward and intricate connections. They are additionally utilized to recognize patterns and clusters within data. In the process of acquiring knowledge, an artificial neural network (ANN) can be constructed with a particular objective, such as the classification of data and the categorization of patterns. Diverse arrangements of neural networks are found in various configurations. Network architectures are formed through the arrangement of neurons and the establishment of connections within and across various layers. It is important to recognize that the objective of cancer prediction diverges slightly from that of cancer detection, although it contributes to the diagnosis and aids in the process of identifying cancer [22]. When it comes to methods for predicting cancer, the following aspects receive particular attention:

i. It involves evaluating the potential risks associated with cancer and predicting an individual's likelihood of developing the disease.

ii. Estimate the likelihood of something happening again in the future and provide a probability for its occurrence.

iii. The final aspects pertain to forecasting the chances of survival.

The remaining part of the paper has been structured into the subsequent parts. Section 2 provides an overview of the latest research discoveries discussed in the literature. Section 3 outlines the methodology that has been proposed for this study. The findings from the ongoing research have been showcased in Section 4. Section 5 provides the final remarks and discusses the potential future directions for this work.

2. Literature Survey

Advancements in molecular biology, genomics, and cutting-edge imaging techniques have significantly enhanced the accuracy of cancer prediction through the collection of molecular-level data. Cancer prediction focuses heavily on understanding one's vulnerability to the disease, the likelihood of its return, and the expected outcome. Various CNN models have been suggested, each with its own emphasis on different aspects of gene expression data modeling, all aimed at predicting the specific type of cancer. In a scholarly paper [23], the researchers employed both guided and self-guided techniques to categorize the specific cancer type. Incorporating backpropagation, supervised learning entails the utilization of a single-layer feedforward neural network for training. Goal is to minimize the error by adjusting the network's parameters. In the self-guided model, non-deterministic and fuzzy-driven c-means clusters were utilized.
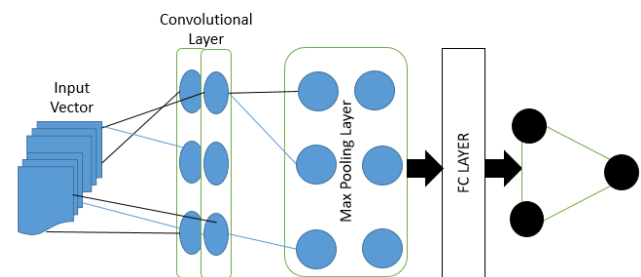
Table 1: Cancer prediction tools and methodology.

Fig 2: Proposed CNN model.

Despite its relatively low accuracy of 76.50%, this approach yields satisfactory outcomes in the initial phases of predictive methods. Diagnosing cancer at the molecular level provides significantly better outcomes compared to traditional approaches, which are often imperfect and can be confusing. In a particular study referenced as [24], a team of scientists employed a neural network to differentiate and group cancer by recognizing the presence of negative correlation. In a particular study referenced as [24], a team of scientists employed a neural network to differentiate and group cancer by recognizing the presence of negative correlation. When utilizing standard reference datasets for prediction, classifiers that exhibit a strong inverse relationship are most effective. An advanced blend of neural

networks utilizes a probabilistic and discrete approach to optimize performance, inspired by the principles of genetic algorithm [25]. The optimization tool driven by PSO aids in the identification of suitable genes and assists in reducing the number of dimensions effectively. Particle swarm optimization (PSO),genetic algorithm belongs to a category of algorithms called collective intelligence. It takes inspiration from the collective intelligence observed in nature and social groups. Particle swarm optimization (PSO) belongs to a category of algorithms called swarm intelligence. It takes inspiration from the collective intelligence observed in nature and social groups. Recent research articles have suggested combining Particle Swarm Optimization (PSO) with local search and path-relinking techniques, showing encouraging outcomes. The particle swarm optimization technique, which is based on PSO, assists in the selection of appropriate genes and contributes to simplifying the dimensions. There have been similar attempts to classify cancer using artificial neural networks (ANNs). When applied to a substantial dataset of B-cell lymphoma, it achieved an 80% accuracy rate. Previous efforts have been undertaken to categorize cancer using artificial neural networks (ANN) [26,27]. CNNs not only achieve high accuracy in classification tasks but also reduce the burden on machine learning experts to manually engineer features by automatically extracting meaningful representations from the data [28]. Since CNN needs a substantial amount of data to gain a good understanding of the problem at hand [29], Harangi [30] explored the potential of employing a series of deep Convolutional Neural Networks (CNNs) to enhance the accuracy of individual models in distinguishing various categories of skin cancer.

new convolutional neural network (CNN) structure consisting of multiple branches for the purpose of classifying skin abnormalities. The structure of the artificial neural network (ANN) was combined with a genetic algorithm that focuses on multiple objectives for the purpose of optimization. The accuracy of the Wisconsin breast cancer database was evaluated by testing it on two distinct tumor categories: malignant and benign tumors [33, 53]. Table 1 presents additional literature reviews and their key findings. Another crucial factor in the detection process is the necessity of having suitable data sets. Additionally, Table 1 displays the specific dataset specifications needed for each approach. Many of these necessitate extensive or sizable datasets for their training and validation, resulting in a higher level of intricacy in the overall system. The main quality of an Artificial Neural Network (ANN) [8, 51] lies in its ability to acquire knowledge. The adjustment process, commonly referred to as "training" or "instruction," enables a computational network to modify its behavior based on the input it receives, resulting in the production of the intended output. Whenever the process of learning is observed, an instructor is present. Having a teacher or supervisor is essential during this type of training to reduce mistakes. The exact desired output values are known for each input pattern. An artificial neural network has the ability to learn and develop new abilities based on the training it receives or the initial experiences it encounters. Once an Artificial Neural Network (ANN) has acquired knowledge during its learning phase, it has the ability to autonomously establish its own structure or organization.
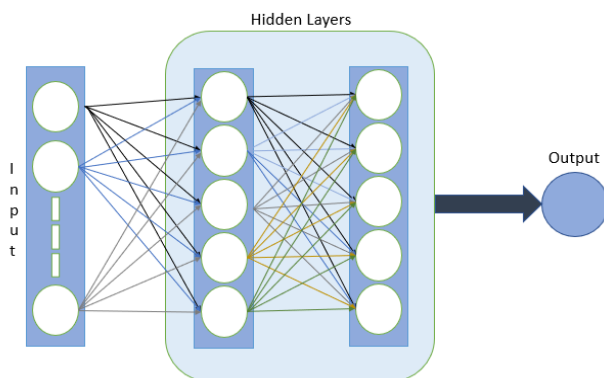

Fig 3: Procedures for neural pattern recognition.

Kawahara et al. demonstrated a linear classifier that utilizes a parameter derived from a pre-trained CNN trained on a dataset comprising 1300 natural images [31] aiming to enhance the accuracy in classifying up to ten different categories of skin abnormalities. The authors introduced a new convolutional neural network (CNN) structure consisting of multiple branches for the purpose of classifying skin abnormalities. Authors of [32] introduced a

Tools/methodology	Year	Accuracy	Dataset requirements	Description
Granular computing [8, 14]	2008	100% (stated)	Large	This algorithm eliminates noise and unwanted genes to predict better
Neural network with MRI image [12, 21]	2010	NA	Large	Neurofuzzy classifiers were used on the brain tumour test data
Support vector machine with fuzzy [22, 28, 31]	2011	92%	Medium	It uses liver cancer datasets for testing. Various micro-ranking–level techniques were implemented to classify
Support vector machine with PSO [23, 34–36]	2012	96%	Medium	Uses breast cancer datasets, but for other datasets, the result and accuracy can deviate
ANN with PSO [24, 37–39]	2012	92.36%	Medium	It was used on the tumour cells. Also implemented on the breast cancer datasets
Particle swarm optimization (PSO) integrated with seeker optimization algorithm (SOA) [25]	2013	~93%	Medium	Liver tumours were analysed and classified
Deep learning [26]	2020	NA	Medium	Using multiomics data for cancer classification

The research article [40, 52] utilized perceptron networks as a key component. This approach was applied to analyze 4026 genes in cases of large B-cell lymphoma. By employing this approach, the identification of cancer achieved a precision rate of 93%, demonstrating high accuracy. While the outcome is satisfactory, when dealing with extensive datasets and intricate values, the level of accuracy may diminish.

3. Proposed Methodology

Various researchers have put forth a range of Convolutional Neural Networks (CNNs) up until now. Each of the models concentrated on distinct factors to simulation of genetic transcription. In earlier research, scientists typically used the sequence of genes as input and aimed to improve the order of these genes to make more accurate predictions. This suggested approach combines CNN with neural pattern recognition to form a unified methodology.

Pattern recognition refers to the skill of recognizing and distinguishing between mathematical symbols, punctuation marks, and printed letters. This model provides a convenient method for structuring the classification issue, although it may lack flexibility and oversimplify the matter. However, it does enable the identification and exploration of numerous notable challenges.
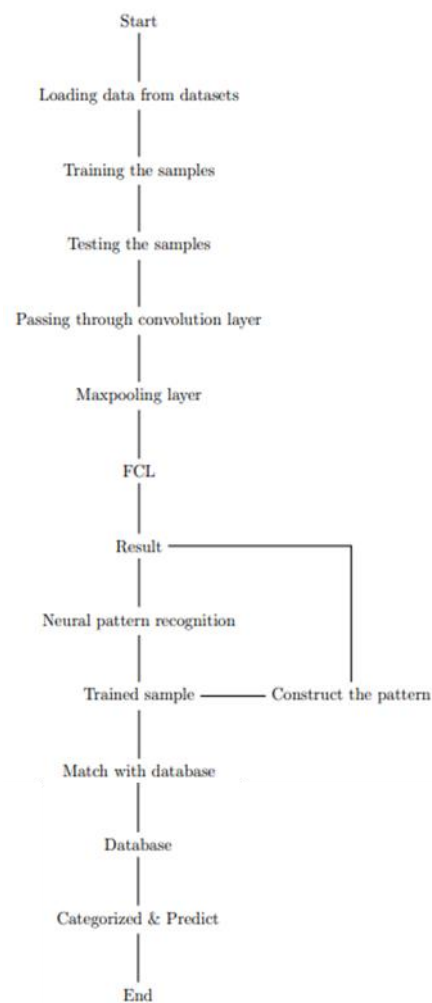

Fig 4: Overall flowchart of the proposed model

Sample size	Kernel	Mean trained score	Mean test score
30	8	0.054	0.197
30	16	0.032	0.167
30	32	0.014	0.132
30	64	0.009	0.121
50	8	0.072	0.196
50	16	0.034	0.154
50	32	0.019	0.141
50	64	0.011	0.132
75	8	0.092	0.189
75	16	0.068	0.163
75	32	0.031	0.145
75	64	0.023	0.137
85	8	0.103	0.178
85	16	0.087	0.156
85	32	0.054	0.143
85	64	0.038	0.129
100	8	0.128	0.172
100	16	0.096	0.154
100	32	0.073	0.142
100	64	0.046	0.121
110	8	0.131	0.167
110	16	0.098	0.154
110	32	0.081	0.136
110	64	0.052	0.120
125	8	0.147	0.159
125	16	0.100	0.141
125	32	0.79	0.129
125	64	0.51	0.117

Table2: Trained and testing parameters for CNN.

Pattern recognition utilizes machine learning techniques to automatically detect and comprehend consistent patterns and regularities within data.

3.1. Proposed Convolution Neural Network: According to studies, CNN models have proven to be highly effective in accurately categorizing various computer vision tasks. These findings suggest that it might be worthwhile to explore the implementation of CNN models for analyzing biological datasets. In this suggested approach, a sole layer of convolution is employed, avoiding the use of multiple layers. The CNN model works with input data represented as directions. And then applies a uni-dimensional kernel operation to process and analyze the input [36,41,42]. First, the genetic information, known as gene expression, is transformed into a numerical representation called a vector. This vector is then processed using a one-dimensional kernel method. The output of this processing is then fed into a max-pooling layer in a convolutional neural network (CNN). This max-pooling layer helps extract the most important features from the processed data. The output from the max-pooling layer is then connected to a fully connected layer (FCL), which helps to further analyze and understand the extracted features. Finally, a prediction layer is used to make predictions or draw conclusions based on the information gathered throughout the entire process as

depicted on Figure 2. Aim of cancer detection is to categorize different types of tumors and discover unique characteristics for each cancer. This enables the development of a smart system that can recognize specific spreading tumor types or detect cancer in its initial stages.
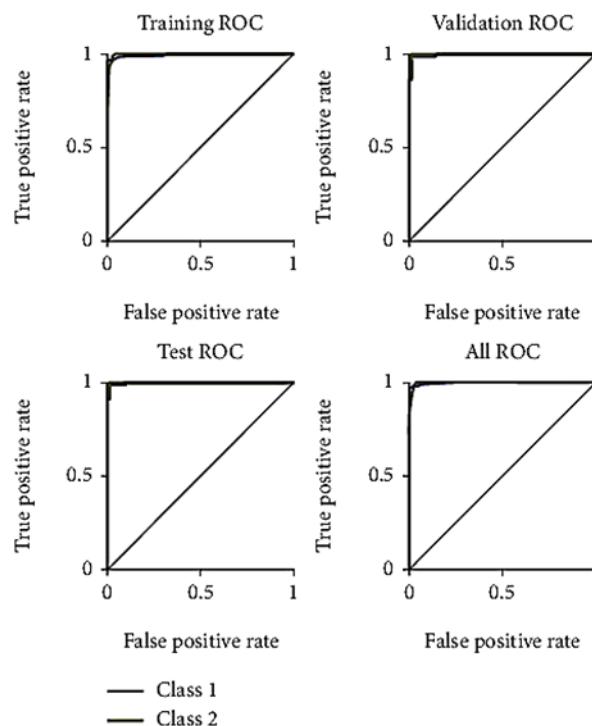

Fig 5: NPR training and test results.

Cancer prediction focuses heavily on determining the likelihood of developing cancer, the chances of cancer returning, and the expected outcome or outlook of the disease. One facet of artificial intelligence referred to as deep learning revolves around the utilization of neural networks featuring multiple layers. This design enhances its capacity for advanced learning and yields more precise outcomes by harnessing cutting-edge techniques such as extensive analysis of vast datasets. Various convolutional neural network (CNN) architectures have been put forth for the purpose of classifying cancer types, with each model highlighting distinct facets of genetic data pertinent to cancer. The input vectors undergo processing in the convolutional layer, followed by the application of the downsampling layer, and finally, they are passed through a fully connected layer. The ultimate result consists of forecasts but with varying set of parameters. In the initial stage, the computation of the gradient of output classes in relation to minute alterations is retained. Every input contributes to the creation of a map that can be utilized for understanding and analysis. The samples were subsequently classified and grouped based on their characteristics. A set of samples were used as input and matched with the dataset features to make predictions. This proposed approach, a genetic indicator with a score surpassing 0.65 is given attention. The selection of the layer size ranged from 30 to

125. Initially, the one-dimensional convolutional neural network (1-D CNN) underwent training using a dataset of 5000 diverse tumor samples. To proceed further, the system was divided, allocating 75% for training purposes and reserving 25% for testing.

3.2. Neural Pattern Recognition (NPR): Pattern recognition involves the utilization of machine learning algorithms to identify consistent patterns and resemblances within datasets. These similarities can now be identified through the examination of past records, the application of statistical techniques, or the utilization of the machine's existing knowledge [36,43,44]. It is necessary to subject the gathered data to filtration and preliminary processing stages. The arrangement is determined prior to selecting the appropriate approach among classification, reversion, and progression to identify and modify based on the numerical type [45,50].
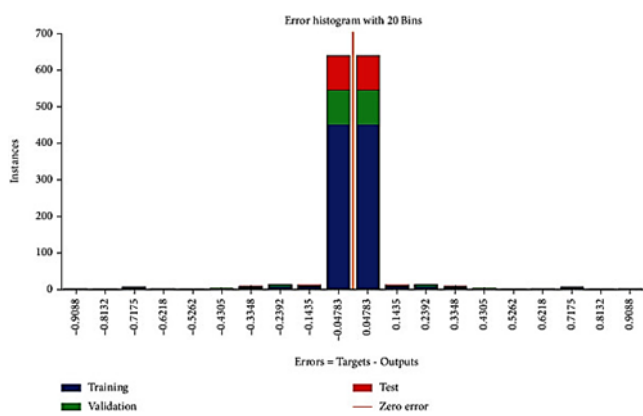

Figure 6: Error histogram

The process of perception can be divided into two fundamental categories: discerning tangible entities and acknowledging conceptual entities. Understanding spatial elements is vital for comprehending tangible objects. Some instances of spatial items include: unique patterns on fingertips, meteorological charts, visual representations, and tangible physical entities. Waveforms and signatures are examples of temporal entities. NPR acknowledging intangible concepts involves recalling a response to an inquiry, a past disagreement or dialogue, and so forth, essentially discerning entities that lack physical existence. Item recognition involves three sequential stages of processing: categorization of features, filtration of input, and identification of the item.

Substantial datasets are necessary to achieve favorable results since the software will consistently produce precise outcomes even with limited training data. There is a possibility that the results may differ when it comes to testing data. A collection of numerous photographs displaying individuals donning masks is necessary when developing a facial recognition system designed specifically for identifying masked faces [46]. The program will utilize the dataset to gather the relevant information. Typically, the training sample accounts for approximately 70 to 80 percent of the entire dataset.

As the precision of the training dataset is on the rise, a portion of the training dataset that the model is not familiar with is selected to determine if the accuracy of that particular subset is also showing improvement [47]. In such a scenario, it becomes crucial for the developer to carefully verify the parameter scores, as failure to do so could necessitate a reassessment of the entire model. Figure 3 depicts the sequential processes involved in identifying the pattern.

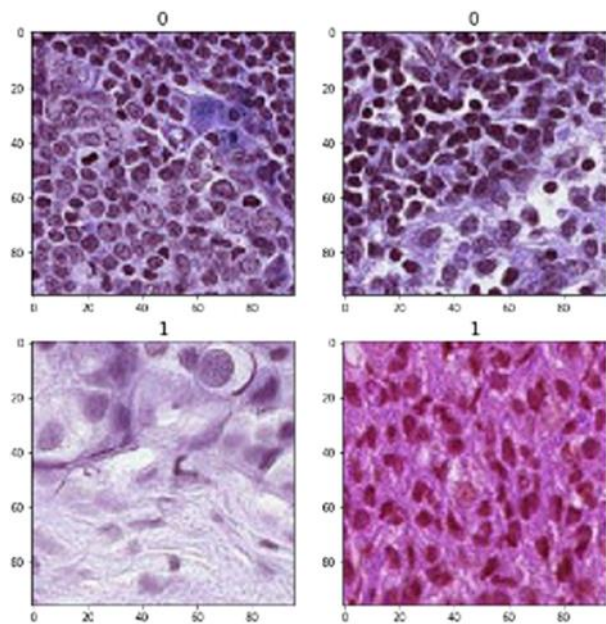The general schematic of the proposed approach is depicted in Figure 4.

Fig 7: Weight diagram for different dataset inputs.

The initial stage involves the utilization of CNN to classify the datasets based on pattern anticipation. In the subsequent phase, NPR employs the identified pattern to further classify the datasets and facilitate pattern prediction.

4. Results

By exclusively training the deep learning machine on tumor samples and subsequently identifying genes associated with cancer, prior research either neglected the tissue origins during the classification of tumor samples or developed two separate models: one focusing solely on transcription factors, and another specifically designed for cancer-related genes (the tumor DL model). The information has been processed using exclusively tumor-related datasets.
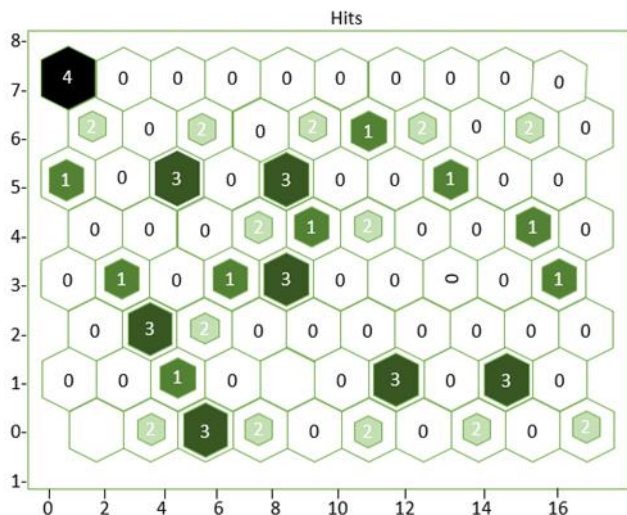
Fig 8: Hit matrix.

The collection consists of data related to breast cancer, kidney tissues, liver tissues, and the digestive systems. Table illustrates the parameters used for training and testing that are transmitted through the one-dimensional Convolutional Neural Network (CNN). Table 2 illustrates the parameters used for training and testing that are transmitted through the one-dimensional Convolutional Neural Network (CNN). During the categorization process, a significant mistake was identified in the database containing information about the liver and digestive tissues. This error occurred due to the unavailability of suitable datasets. The effectiveness of a classifier model is assessed using a measurement known as the ROC curve, which is alternatively termed the curve of receiver operating characteristics. ROC curve emphasizes the ability of the classifier model to correctly identify positive instances by depicting the proportion of true positives relative to false positives. The ROC curve is a useful tool for evaluating the effectiveness of classification methods at different threshold levels. It provides a measure of efficiency in distinguishing between classes. The AUC (Area Under the Curve) serves as a quantification of how well separable a given measure is, while the ROC (Receiver Operating Characteristic) denotes a graphical representation of the probability. The area under the curve of the receiver operating characteristic (ROC) curve offers valuable information regarding the benefits of utilizing the test to address the fundamental question. ROC curves are commonly employed to visually depict the interplay and compromise between sensitivity and specificity. Figure 5 displays the outcomes of the NPR study, showcasing highly favorable results that have been confirmed and verified using the ROC method. Indeed, the figure has depicted varying rates for both true positives and false positives. Through the evaluation of the proportion of accurate positive predictions relative to incorrect positive predictions, the ROC curve demonstrates the classifier model's ability to differentiate between classes. The ROC curve serves as a performance indicator for classification

tasks across various threshold levels, gauging the effectiveness of the model. AUC signifies the degree or assessment of distinguishability, while ROC is a curve that illustrates probabilities. This demonstrates the model's ability to efficiently discern among various classifications. The advantages of employing the examination to address the real inquiry become apparent through the calculation of the area beneath the ROC curve. The histogram depicting the errors in the results accurately represents the values. Figure 6 displays the errors. Figure 7 displays the respective inputs labeled as 1, 2, 3, and 4, representing the weight values associated with various datasets, namely liver, kidney, breast, and digestive. The hit matrix (Figure 8) provides a clear indication of the most favorable probability for accurately forecasting the pattern value. In this situation, the highest possible value falls within the range of 9 and 10, specifically at 4, with a precision level of 94%. The connection between weight 1 and weight 2 can be observed in the diagram presented in Figure 9. This suggests that the weights have a crucial role in the neural network system, helping it recognize the connections between various parameters, as depicted in Figure 9. Now, different sets of data are analyzed to assess the performance of the system and determine its accuracy. Table 3 presents the results obtained with various datasets.

Cancer is widely recognized as one of the most devastating illnesses. Early detection of cancer is crucial for effective treatment. The aim of achieving early cancer detection is to swiftly identify individuals experiencing symptoms, enabling them to receive timely treatment and enhance their chances of successful recovery. By providing timely healthcare services, detecting cancer at an early stage leads to better treatment results, making it an essential strategy for public health in every situation. Prompt diagnosis is one strategy, whereas assessment is another approach. This refers to the potential identification of undiscovered illnesses among a group of individuals who seem to be in good health and without any apparent symptoms. This can be achieved through affordable and efficient tests, evaluations, or other methods applied to the specific population.

Cancer data consists of numerous or countless genes. The expression level of genes is assessed using a DNA microarray. Analyzing microarray gene expression data can be quite difficult due to its distinctive and overwhelming characteristics. Identifying appropriate genes from a vast pool of millions of genes is a challenging task.
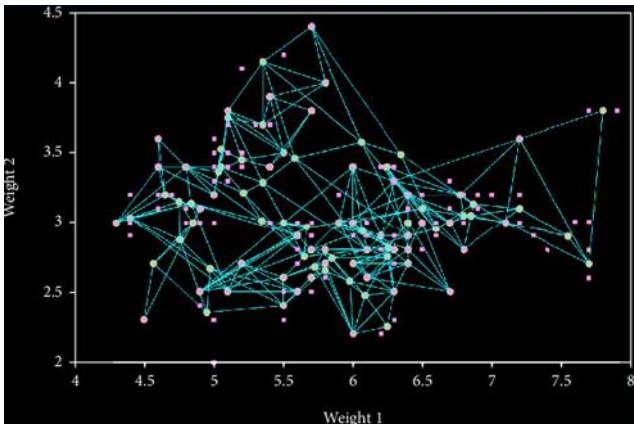
Fig 9: Relation between different parameters with weights.

Table 3: Accuracy with different datasets.

Dataset	Trial	Accuracy
Kidney Cancer	20	96.01%
Breast Cancer	20	95.1%
Liver tissues	20	93.2%
Liver Cancer	20	92.4%
Mixed data	50	94.01%

Researchers are utilizing machine learning and statistical techniques such as decision trees, k-nearest neighbors, support vector machines, and neural networks to precisely examine tumors. These advanced techniques aid in effectively classifying and understanding tumors [48]. Numerous researchers have shown a keen inclination towards employing neural networks for the purpose of categorizing diverse types of cancer cells [49]. Most neural networks are highly effective at correctly categorizing cancerous cells. This suggested approach effectively analyzes the information and accurately forecasts cancer occurrences, achieving a 94% precision rate. Due to the progress made in protein, genetic, and advanced imaging techniques, it is now feasible to predict cancer with greater accuracy by collecting detailed molecular information. Many of the conventional methods employed in medical environments to predict cancer are riddled with shortcomings and create ambiguity, whereas the utilization of micro gene-level diagnostics leads to significantly improved outcomes. Based on the study, CNN models are utilized to achieve more accurate categorizations in different visual analysis tasks. Therefore, it is logical to try and use them on biological datasets. Early detection plays a vital role in ensuring the most effective treatment for cancer. Cancer is composed of a vast number of genes, ranging from thousands to millions, that hold crucial information about the disease. A DNA microarray is employed for measuring the levels of genetic markers.

5. Conclusion

Several researchers have integrated neural network methodologies with optimization algorithms like Particle Swarm Optimization (PSO) in order to further enhance accuracy. These techniques are employed to reduce the number of dimensions, narrow down the search area, and consequently shorten the time required for training neural networks. FLANN by itself achieves an accuracy rate of 63.4 percent, whereas when combined with PSO, the categorization rate increases to 92.36 percent. To enhance the reliability of our cancer predictions and assist in their comprehension, our study encompassed several noteworthy subjects. To explore an effective method for identifying cancer types using unstructured gene-based expressions, various convolutional neural network (CNN) designs were suggested. We have developed a unique CNN-NPR structure that utilizes complex gene expression data to accurately determine the type of cancer, considering the tissue it originates from. The specimens encompass information concerning gastrointestinal systems, kidney tissues, liver tissues, and tissues from breast cancer. ROC offers and confirms highly favorable results. Identifying important genes from the vast pool of millions of genes can be a daunting task. Machine learning models like classification trees, k-nearest neighbours, support vector machines, and artificial neural networks are being utilized to accurately depict tumors. The suggested approach, characterized by a streamlined CNN-NPR design and reduced reliance on tissue origin compared to previous research, data's accuracy in predicting is at a commendable 94%. We can utilize the CNN-NPR framework to elucidate distinct cancer signals for each cancer type, with the objective of further improving and refining it in the times ahead. The ultimate objective is to identify markers that can aid in the early detection of cancer. To enhance the precision of neural networks in future research, it is advisable to augment the quantity of intermediate neurons within the hidden layer. There are various techniques that can be employed to enhance the efficiency of the classifier during the training process of an Artificial Neural Network (ANN).

Data Availability

The corresponding author can provide the data used in this study upon receiving a request.

Conflicts of Interest

The authors assert that they have no competing interests..

References

[1] Y. Wang, F. S. Makedon, J. C. Ford, and J. Pearlman, "HykGene: a hybrid approach for selecting marker genes for phenotype classification using microarray gene expression data," Bioinformatics, vol. 21, no. 8, pp. 1530–1537, 2005.

[2] T. Thakur, I. Batra, M. Luthra et al., "Gene expression-assisted cancer prediction techniques," Journal of Healthcare Engineering, vol. 2021, Article ID 4242646, 9 pages, 2021.

[3] S. Kaur and G. Kaur, "Threat and vulnerability analysis of cloud platform: a user perspective," in 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom), pp. 533–539, New Delhi, India, 2021.

[4] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," Computational and Structural Biotechnology Journal, vol. 13, pp. 8–17, 2015.

[5] Y. Wang, I. V. Tetko, M. A. Hall et al., "Gene selection from microarray data for cancer classification--a machine learning approach," Computational Biology and Chemistry, vol. 29, no. 1, pp. 37–46, 2005.

[6] G. Jindal and G. Kaur, "A comprehensive overview of quality enhancement approach-based biometric fusion system using artificial intelligence techniques," Communication and Intelligent Systems, vol. 204, pp. 81–98, 2021.

[7] S. Chaudhury, N. Shelke, K. Sau, B. Prasanalakshmi, and M. Shabaz, "A novel approach to classifying breast cancer histopathology biopsy images using bilateral knowledge distillation and label smoothing regularization," Computational and Mathematical Methods in Medicine, vol. 2021, Article ID 4019358, 11 pages, 2021.

[8] S.-B. Cho and H.-H. Won, "Cancer classification using ensemble of neural networks with multiple significant gene subsets," Applied Intelligence, vol. 26, no. 3, pp. 243–250, 2007.

[9] Y. Hu, K. Ashenayi, R. Veltri, G. O'Dowd, G. Miller, and H. R. Bonner, "A comparison of neural network and fuzzy c-means methods in bladder cancer cell classification," in Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94), pp. 3461–3466, Orlando, FL, USA, 1994.

[10] P. H. Levine, K. Ajmera, B. O'Neill, V. Venkatesh, P. Garcia-Gonzalez, and H. J. Hoffman, "Demographic factors related to young age at diagnosis of chronic myeloid leukemia in India," Clinical Epidemiology and Global Health, vol. 4, no. 4, pp. 188–192, 2016.

[11] A. Elfiky, M. J. Pany, R. B. Parikh, and Z. Obermeyer, "Development and application of a machine learning approach to assess short-term mortality risk among patients with cancer starting chemotherapy," JAMA Network Open, vol. 1, no. 3, article e180926, 2018.

[12] S. L. Bangare, G. Pradeepini, and S. T. Patil, "Implementation for brain tumor detection and three dimensional visualization model development for reconstruction," ARPN Journal of Engineering and Applied Sciences (ARPN JEAS), vol. 13, no. 2, pp. 467–473, 2018.

[13] K. A. Tran, O. Kondrashova, A. Bradley, E. D. Williams, J. V. Pearson, and N. Waddell, "Deep learning in cancer diagnosis, prognosis and treatment selection," Genome Medicine, vol. 13, no. 1, 2021.

[14] Y. Tang, Y.-Q. Zhang, Z. Huang, H. Xiaohua, and Y. Zhao, "Recursive fuzzy granulation for gene subsets extraction and cancer classification," IEEE Transactions on Information Technology in Biomedicine, vol. 12, no. 6, pp. 723–730, 2008.

[15] S. Akanksha and K. Parminder, "Optimized liver tumor detection and segmentation using neural network," International Journal of Recent Technologyand Engineering(IJRTE), vol. 2, no. 5, pp. 7–10, 2013.

[16] H.-H. Won and S.-B. Cho, "Paired neural network with negatively correlated features for cancer classification in DNA gene expression profiles," in Proceedings of the International Joint Conference on Neural Networks, 2003, pp. 1708–1713, Portland, OR, USA, 2003.

[17] R. Xu, X. Cai, C. Donald, and I. I. Wunsch, "Gene expression data for DLBCL cancer survival prediction with a combination of machine learning technologies," in 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, pp. 894–897, Shanghai, China, 2005.

[18] V. Bevilacqua, G. Mastronardi, F. Menolascina, P. Pannarale, and A. Pedone, "A Novel Multi-Objective Genetic Algorithm Approach to Artificial Neural Network Topology Optimisation: The Breast Cancer Classification Problem," in The 2006 IEEE International Joint Conference on Neural Network Proceedings, pp. 1958–1965, Vancouver, BC, Canada, 2006.

[19] L. Ziaei, A. R. Mehri, and M. Salehi, "Application of artificial neural networks in cancer classification and diagnosis prediction of a subtype of lymphoma based on gene expression profile," Journal of Research in Medical Sciences, vol. 11, no. 1, pp. 13–17, 2006.

[20] H. Takahashi, Y. Murase, T. Kobayashi, and H. Honda, "New cancer diagnosis modeling using boosting and projective adaptive resonance theory with improved reliable index," Biochemical Engineering Journal, vol. 33, no. 2, pp. 100–109, 2007.

[21] D. M. Joshi, N. K. Rana, and V. M. Misra,

"Classification of brain cancer using artificial neural network," in 2010 2nd International Conference on Electronic Computer Technology, pp. 112–116, Kuala Lumpur, Malaysia, 2010.

[22] P. Rajeswari and G. S. Reena, "Human liver cancer classification using microarray gene expression data," Proceedings of International Journal of Computer Applications, vol. 34, no. 6, pp. 25–37, 2011.

[23] B. Sahu and D. Mishra, "A novel feature selection algorithm using particle swarm optimization for cancer microarray data," Procedia Engineering, vol. 38, pp. 27–31, 2012.

[24] S. Swathi, G. Anjan Babu, R. Sendhilkumar, and S. N. Bhukya, "Performance of ART1 network in the detection of breast cancer," Proceedings of International conference on Computer design and Engineering (ICCDE 2012), vol. 49, pp. 100–105, 2012.

[25] M.S Kumar, "Prediction of Heart Attack from Medical Records Using Big Data Mining", International Journal of Intelligent Systems and Applications in Engineering, 2023, 11(4s), pp. 90–99.

[26] M.S Kumar, "Nature-Inspired Optimisation-Based Regression Based Regression to Study the Scope of Professional Growth in Small and Medium Enterprises, International Journal of Intelligent Systems and Applications in Engineering, 2023, 11(4s), pp. 100–108.

[27] Natarajan, V.A., "Improving QoS in Wireless Sensor Network routing using Machine Learning Techniques", Proceedings of the 1st IEEE International Conference on Networking and Communications 2023, ICNWC 2023, 2023.

[28] M.S Kumar, "Chronic Kidney Disease Prediction Using Machine Learning", Journal of Advances in Information Technology, 2023, 14(2), pp. 384–391.

[29] D Ganesh, "Deep Convolution Neural Network Based solution for Detecting Plant Diseases", Journal of Pharmaceutical Negative Results, 2022, 13, pp. 464–471

[30] D Ganesh, "Implementation of Novel Machine Learning Methods for Analysis and Detection of Fake Reviews in Social Media", 2nd International Conference on Sustainable Computing and Data Communication Systems, ICSCDS 2023 - Proceedings, 2023, pp. 243–250

[31] Ganesh, D., Kumar, T. P., & Kumar, M. S. (2021). Optimised Levenshtein centroid cross-layer defence for multi-hop cognitive radio networks. IET Communications, 15(2), 245-256.

[32] Sushama et.al, "Impact of COVID-19 pandemic and the diagnosis of the virus in the human body",World Journal of Engineeringthis link is disabled, 2022, 19(5), pp. 652–657

[33] Anantha Natarajan, V., Kumar, M. S., & Tamizhazhagan, V. (2020). Forecasting of Wind Power using LSTM Recurrent Neural Network. Journal of Green Engineering, 10.

[34] Sushama et.al, Automated extraction of non-functional requirements from text files: A supervised learning approach", Handbook of Intelligent Computing and Optimization for Sustainable Development, 2022, pp. 149–170D.

[35] A. Omondiagbe, S. Veeramani, and A. S. Sidhu, "Machine learning classification techniques for breast cancer diagnosis," IOP Conference Series: Materials Science and Engineering, vol. 495, article 012033, 2019.

[36] S Girinath,"Deep Learning-based Segmentation and Computer Vision-based Ultrasound Imagery Techniques", 2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT),DOI: 10.1109/ICECCT56650.2023.10179723, ISBN:978-1-6654-9361-1, 2023.

[37] Sushama, C.," Global Identification Passport: A Unique Cloud based Passport Model", Proceedings - 5th International Conference on Smart Systems and Inventive Technology, ICSSIT 2023, 2023, pp. 801–805

[38] M. S. Kumar, "AI-driven cybersecurity modeling using quantum computing for mitigation of attacks in IOT-SDN network", Quantum-Safe Cryptography Algorithms and Approaches: Impacts of Quantum Computing on Cybersecur, 2023, pp. 37–47.

[39] M. S. Kumar, "Power aware virtual machine placement in IaaS cloud using discrete firefly algorithm", Applied Nanoscience (Switzerland)this link is disabled, 2023, 13(3), pp. 2003–2011.

[40] Natarajan, V.A., "Improving QoS in Wireless Sensor Network routing using Machine Learning Techniques", Proceedings of the 1st IEEE International Conference on Networking and Communications 2023, ICNWC 2023, 2023T. O. Nielsen, R. B. West, S. C. Linn et al., "Molecular characterisation of soft tissue tumours: a gene expression study," Lancet, vol. 359, no. 9314, pp. 1301–1307, 2002.

[41] K. Jyothi Prakash. "Internet of things: IETF protocols, algorithms and applications." Int. J. Innov. Technol. Explor. Eng 8.11 (2019): 2853-2857.

[42] Sangamithra, B. "A memetic algorithm for multi objective vehicle routing problem with time windows." 2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE). IEEE, 2017.

[43] P. S. Nelson, N. Clegg, H. Arnold et al., "The program of androgen-responsive genes in neoplastic prostate epithelium," Proceedings of the National Academy of Sciences of the United States of America, vol. 99, no. 18, pp. 11890–11895, 2002.

[44] Y. Guo, X. Shang, and Z. Li, "Identification of cancer subtypes by integrating multiple types of transcriptomics data with deep learning in breast cancer," Neurocomputing, vol. 324, pp. 20–30, 2019.

[45] S. Chatterjee, A. Iyer, S. Avva, A. Kollara, and M. Sankarasubbu, "Convolutional neural networks in classifying cancer through DNA methylation," 2018, https://arxiv.org/abs/1807.09617.

[46] Z. Si, H. Yu, and Z. Ma, "Learning deep features for DNA methylation data analysis," IEEE Access, vol. 4, pp. 2732–2737, 2016.

[47] Y. C. Chiu, H. H. Chen, T. Zhang et al., "Predicting drug response of tumors from integrated genomic profiles by deep neural networks," BMC Medical Genomics, vol. 12, Supplement 1, p. 18, 2019.

[48] P. Luo, Y. Ding, X. Lei, and F. X. Wu, "deepDriver: predicting cancer driver genes based on somatic mutations using deep convolutional neural networks," Frontiers in Genetics, vol. 10, p. 13, 2019.

[49] Y. C. Chiu, H. I. H. Chen, A. Gorthi et al., "Deep learning of pharmacogenomics resources: moving towards precision oncology," Briefings in Bioinformatics, vol. 21, no. 6, pp. 2066–2083, 2019.

[50] S. Depuru, K. Santhi, K. Amala, M. Sakthivel, S. Sivanantham and V. Akshaya, "Deep Learning-based Malware Classification Methodology of Comprehensive Study," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 322-328, doi: 10.1109/ICSCDS56580.2023.10105027.

[51] S. Depuru, S. Sivanantham, K. Amala, V. Akshaya, M. Sakthivel and P. S. Kusuma, "Empirical Study of Human Facial Emotion Recognition: A Deep Learning," *2023 International Conference on Computer Communication and Informatics (ICCCI)*, Coimbatore, India, 2023, pp. 1-6, doi: 10.1109/ICCCI56745.2023.10128492.

[52] Sivakumar Depuru , Anjana Nandam , P.A. Ramesh , M. Saktivel , K. Amala , Sivanantham, "Human Emotion Recognition System Using Deep Learning Technique", Journal of Pharmaceutical Negative Results, vol. 13, no. 4, pp. 1031–1035, Nov. 2022.

[53] K. Pujitha , Kattamanchi Prem Krishna , K. Amala , Annavarapu Yasaswini , Sivakumar Depuru , Kopparam Runvika, "Development of Secured Online Parking Spaces", Journal of Pharmaceutical Negative Results, vol. 13, no. 4, pp. 1010–1013, Nov. 2022.

[54] Kulkarni, L. . (2022). High Resolution Palmprint Recognition System Using Multiple Features. Research Journal of Computer Systems and Engineering, 3(1), 07–13. Retrieved from https://technicaljournals.org/RJCSE/index.php/journal/article/view/35

[55] G, M. ., Deshmukh, P. ., N. L., U. K. ., Macedo, V. D. J. ., K B, V. ., N, A. P. ., & Tiwari, A. K. . (2023). Resource Allocation Energy Efficient Algorithm for H-CRAN in 5G. International Journal on Recent and Innovation Trends in Computing and Communication, 11(3s), 118–126. https://doi.org/10.17762/ijritcc.v11i3s.6172