International Journal of Basic and Applied Sciences, 14 (5) (2025) 617-624

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/gz68s920
Research paper

Classification and Prediction of Pleuro Pulmonary Blastoma Using Deep Learning Models

Arigela Jyothi [1], Janapareddy Uttara Alekhya [2], Anuradha Patnala [3], Mosuri Santhosh Kumar [4 *], Gorli Ramya [5], Rajendra Kumar Ganiya [6]

[1] Department of Computer Science and Engineering, Keshav Memorial Institute of Technology (KMIT), Narayanguda, Hyderabad, Telangana, India
[2] Department of Advanced Computer Science and Engineering, Vignan's Institute of Information Technology (A), Visakhapatnam, Andhra Pradesh, India
[3] Department of Computer Science and Engineering, Centurion University of Technology and Management, Vizianagaram, Andhra Pradesh, India
[4] Department of Electrical and Electronics Engineering, GMR Institute of Technology (A), Rajam, Andhra Pradesh, India
[5] Department of Computer Science and Engineering, Vignan's Institute of Engineering for Women (A), Visakhapatnam, Andhra Pradesh, India
[6] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India
**Corresponding author E-mail: msanthosh.gmrit@gmail.com*

Received: June 29, 2025, Accepted: September 10, 2025, Published: September 17, 2025

Abstract

Pleuro-pulmonary Blastoma (PPB) is an uncommon childhood illness caused by embryonic malignancies that have aberrant tissue growth on the pleural surfaces and in the lung parenchyma. Tumors can develop by several methods, allowing them to be categorized into categories I, II, and III. Each type is associated with distinct variances in the age at which they are diagnosed and the prognosis. Our goal was to provide a comprehensive examination of the relevant literature, outlining the features of these tumors and the use of multidisciplinary approaches to treat them, with a specific emphasis on surgical intervention. As described in prior writings, pathologists use patient samples such as MRI scan images, DNA sequencing, etc., to predict and categorize PPB variations. Various machine learning methods are utilized to extract features from proteins according to their protein family, but in the suggested methodology, we used various deep learning models to automatically extract features and for classification. We considered using Deep Learning algorithms, such as ResNet-based Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM), by taking into consideration protein sequence as input data to obtain accurate results and perform the prognosis of type I, II, and III. We discover the connection between the functional annotations of unaligned amino acid sequences and these deep learning models.

Keywords: *Pleuro Pulmonary Blastoma, Machine Learning, Classification, ResNet-based Convolutional Neural Network, Long Short-Term Memory*

1. Introduction

Worldwide, cancer kills more people than any other disease today. The key to a favourable prognosis is being diagnosed early. Liver, lung, prostate, cervical, and breast cancers are only a few of the many types of this disease. According to Global Cancer Statistics 2020, there are around 36 distinct cancers that affect people across the globe. While there are various cancers, lung cancer is by far the most prevalent. As of 2020, it is expected to have a death rate of 1.8 million, or 18 percent, making it a leading cause of death [1]. The primary focus of this essay is lung cancer and its variations. There are essentially two types of lung cancer variations that can exist. These types of lung cancer are prevalent in children, and among them is non-small cell lung cancer, which is also known as pleural pulmonary blastoma PPB in children. It is a type of cancer that typically affects children and develops in the chest. The term used to describe the presence of a condition in the lungs or their covers is pleura. Types I, II, and III of pleuropulmonary blastoma are recognized. Symptoms of pulmonary blastoma type 1 (PPB) include the growth of cysts in the lungs in infants and young children [2]. There are two types of pleuropulmonary blastomas (PPBs). Type II PPBs have both solid and cystic components, while Type III PPBs are solid neoplasms only. Type II and Type III PPB are common in children older than 2 years old. Another type is Type I-r, characterized by regression or regressing, and it is found in PPB. The Type I-r will have a similar appearance to Type I PPB. Unlike Type I, Type I-r does not exhibit malignant cells. As previously

stated, the occurrence of Type I, II, and III PPB is more prevalent in children below the age of eight. PPB can occasionally impact individuals of both adult and adolescent age groups. Type I PPB can occur in individuals of all age groups.
Type I PPB cases in youngsters may be overlooked when comparing them to Type II and Type III PPB. Approximately 89 percent of Type I cases can be cured; however Type II and Type III PPB may continue to occur. Recent studies indicate that between 50 and 70 percent of children with these malignancies have achieved a cure. Surgery or chemotherapy is used to treat Type I tumors, while Type II and Type III cancers are treated with radiation therapy in addition to other treatments [3,4].
The following symptoms are frequently observed in children with PPB, according to the National Organization for Rare Disorders: 1) Respiratory Distress or Breathing Difficulties (dyspnea), 2) Pneumonia/General disease symptoms include fever, headache, fatigue, coughing, and difficulty breathing. DICER1 gene variation in the person or family is the primary underlying cause of PPB [5,6].
Early cancer detection is necessary to reduce patient risk; first, we must forecast the type of cancer and take appropriate action. We introduced Deep Learning Models for PPB prediction and classification. We used deep learning models rather than machine learning algorithms in the proposed methodology because machine learning algorithms required us to manually extract features from proteins for categorization, whereas deep learning models do so automatically. In this study, we focus on deep learning models that use protein sequences from PPB patients' genome data to categorize data.

2. Literature Survey

Sun, Y et al [7], using deep learning in genomes, these techniques allow us to discover connections between genetic variants and phenotypes. All of the genomic variants are considered without regard to segregating locations. This technique makes use of biological neural networks, which are then used to build new deep neural network models for recognizing cancer through a large number of mutations. The GDL approach combines model training and data processing. They used 6083 datasets, 12 cancer type variants from diverse samples, and 1991 healthy individuals from the 1000 Genomes Project database. Mutect2, a Bayesian classifier, detects somatic mutations. They illustrated with a Tensor Flow Deep Neural Network model. In this multi-class classification issue, accuracy, loss, validation accuracy, and validation loss are calculated from all validation subset data.
Song, Y [8], explored artificial intelligence (AI) models to predict the presence of EGFR and KRAS mutations using CT characteristics. For non-small cell lung tumors, these two models serve as genetic markers. Shape, statistics, pathology, and deep learning were extracted from 144 tumor CT scans. Using data from ResNet-34 feature extraction, logistic regression, support vector machine, and random forest were trained to enhance deep learning. Performance evaluation uses confusion matrices, 10-fold cross-validation, and ROC curve area. Logistic regression and support vector machines perform well and have accuracy rates of 85% and 84% for predicting EGFR mutations, respectively. SVM exhibits the best results and has an accuracy of roughly 73% for KRAS mutation prediction.
Adetiba, E [9], compares the experimental results for both the ensemble and non-ensemble types of lung cancer between SVM and the artificial neural network. For training, they use nucleotide and mutation samples from patients. ANN ensemble and HOG produce the best results, with an accuracy rate of 95.90 percent and a mean square error of 0.0159. Ensemble learning, which involves merging numerous deep CNN learners, was employed for pulmonary nodule classification [10]. The LIDC-IDR1 database is used to extract 743 nodules for this purpose using CT image patches. Using 10-fold cross-validation, they trained and evaluated eight deep CNN learners with different architectural designs. By utilizing ensemble learners, we are able to achieve an impressive prediction accuracy of 84%.
Shaikh, F [11] assesses the efficacy of machine learning and deep learning methodologies in modeling the evolution of cancer in their research study. The participants engaged in a comprehensive discussion on different machine learning techniques, highlighting their strengths and weaknesses in terms of forecasting various types of cancer. They presented multiple case examples, each showcasing the precision of a certain machine learning classifier. The study discovered that Artificial Neural Networks (ANN) are extensively utilized and apply several alternative learning methodologies to predict diverse types of cancer. Although the outcome was satisfactory, there is room for improvement in the preparation of the experiment to ensure consistent data quality and quantity.
As mentioned above, researchers apply a variety of ML and DL algorithms for the categorization and prediction of lung cancer, pulmonary nodules, and other types of cancer. Pleura Pulmonary Blastoma prediction and categorization based on the information gathered from the Genome Sequence was the main focus of this paper.

3. Methodology

3.1 Data collection

Dicer1.xlsx, which contains attributes such as variation, protein sequence, and class, DICER family, was used to compile the data from 927 patients for the prediction and categorization of PPB variants. Table 1 [12] contains the dataset's metadata information.

Table 1: Metadata Information of the Dataset

Column name	Type
Variant	String
Protein	Sequence
Family	String
Class	Number

3.2 Data Preprocessing

The most important phase in a machine learning or deep learning model, if the data consists of categorical rather than numerical data, is data preparation. The 20 amino acid protein sequence cannot be directly input into the deep learning model, so I used one-hot encoding, which replaces the protein sequence code with numerical values such as A-1, C-2, D-3, E-4, G-5, H-6, etc. If the code is not found in the dictionary, it replaces it with 0 [13].
Based on the protein sequence obtained, several ML techniques are used to predict and categorize pleural pulmonary blastoma [14]. Originally, we used the dataset, which is a collection of .fasta sequences. A single line describing the sequence data is followed by a certain number of lines in the .fasta format. The first character in this format is greater than the symbol. These .fasta files are used as input in a hot encoding technique to extract K-mer counts or n-grams; using these k-mer counts, we were able to identify several protein variations and

classify their sequences [15]. These protein sequences are utilized as input to the machine learning algorithm, which uses them to classify data. The procedure described above is depicted in Fig. 1 using a flowchart. Since extracting characteristics from data is the fundamental challenge for machine learning algorithms, we used a variety of deep learning models to predict the PPB here [16].

3.3 Deep learning Classification Models

We presented two deep learning models—Bidirectional LSTM and ResNet-based CNN—to circumvent bottlenecks produced by machine learning algorithms [17, 18,19]. Basically, twenty different types of amino acids make up protein sequences; these amino acids determine the structure and function of these protein sequences. Therefore, I need to comprehend the relationship between amino acids and protein sequence for categorization to anticipate pleural pulmonary blastoma. The proposed methodology's process flow is depicted in Fig. 2 [20].

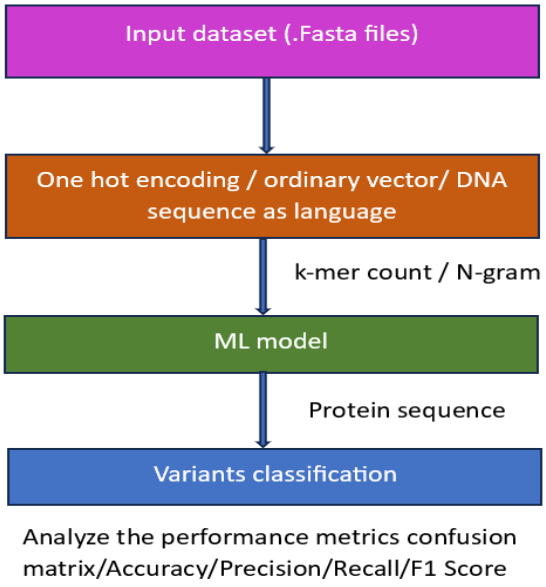

Fig.1: Steps in Pre-processing

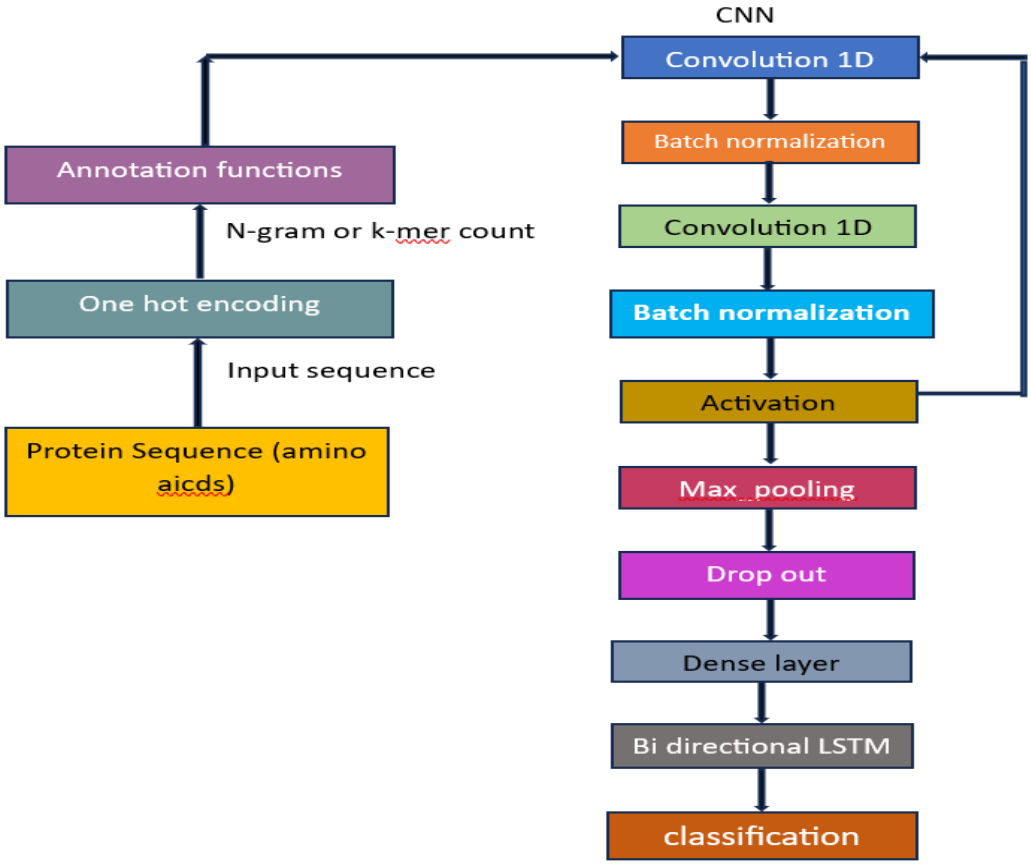

Fig.2: Process flow of Proposed Methodology

The detailed description and model architecture of these algorithms are given below

3.4 Bi-Directional LSTM

By feeding the output of one state into the input of the next, this type of recurrent neural network can retain information about the input sequence. When the input is a shorter sequence, an RNN is frequently utilized. Because they have the unique ability to retain only a subset of prior knowledge for an extended length of time, these LSTM neural networks are updated or improved versions of RNNs. They can remember the desired information and delete the undesired information by using this function. A bi-directional long short-term memory takes in data in two directions: first, from the past, and second, from the future. Fig. 3 [21] depicts the Bidirectional LSTM architecture.

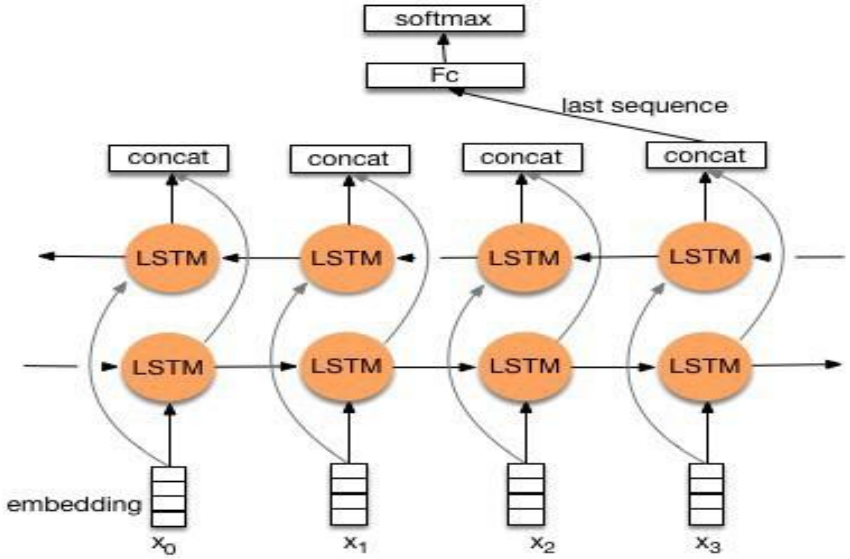

Fig.3: Bidirectional LSTM architecture

This model employs two varieties of RNN, one for backward selection and the other for forward selection. This model's initial layer employs an embedding layer to transform words into 64-by-128-pixel vectors, following which a bidirectional LSTM is used with 128-by-128-pixel filters, and finally, the output is handed to a dense layer for output flattening. The categorization in this model is done using the softmax function. The suggested bidirectional LSTM's full design is shown in Table 2 [22, 23].

Table 2: Complete Architecture of Proposed Resnet-Based CNN

Layer1 (type1)	Output1 Shape1	Param1 #
Input1_1 (Input Layer)	[(None, 40)]	0
Embedding1 (Embedding1)	(None, 40, 64)	1408
Bidirectional (Bidirectional)	(None, 128)	66048
Dropout1 (Dropout1)	(None, 128)	0
Dense1 (Dense1)	(None, 5)	645

ResNet-based CNN: It repeatedly employs CNN blocks and residual blocks. To avoid the vanishing gradient problem with ResNet-based deep neural networks, the challenge is to train the data. By using a skip connection or shortcut connection and adding initial input to the output of the block of the convolution layer, CNN introduces blocks of residuals. Fig. 4 shows a single residual block [24].

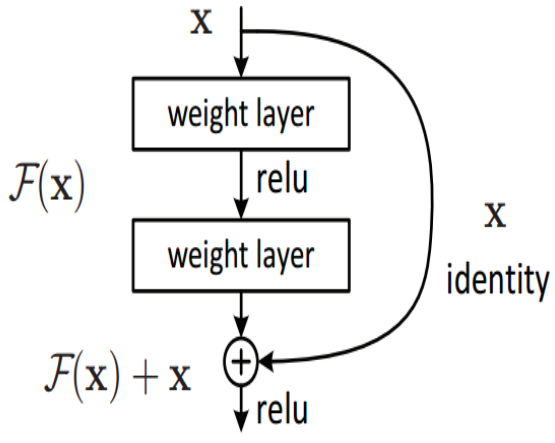

Fig.4: Single Residual block

We must extract the features from the dataset using a ResNet-based CNN, which uses blocks of CNN layers. This model makes use of hyperparameters like the kernel size and the number of filters. The suggested ResNet-based CNN's whole architecture is displayed in Table 3 [25]. The words are transformed into a vector using the first layer in this suggested model, which has a 64-dimensional embedding layer. The model then uses 4 convolution layers with filters of 128 and a kernel size of 2X2 to extract features using a ReLU activation function.

The model applies batch normalization following each convolution and utilizes residual blocks of CNN with filters of 128 [26]. Incorporating a 2x2 max pooling layer reduces the feature map dimensions, which are subsequently flattened into a single column vector. The layer dense with neurons of sizes 128 and 64 receives the output last. For classification purposes, this model employs the multi-class classification function known as softmax [27].

Table 3: Complete Architecture of Proposed Resnet-Based CNN

Layer1 (type)	Output1 Shape1	Param1 #	Connected by
Input1_2 (Input Layer)	(None, 40, 22)	0	[]
conv1d1 (Conv1D1)	(None, 40, 128)	2944	['input1_2 [0] [0]']
Batch normalization1 (Batch Normalization1)	(None, 40, 128)	512	[‘conv1d1 [0][0]']
Activation1 (Activation1)	(None, 40, 128)	0	['batch-normalization1’]
conv1d1_1 (Conv1D)	(None, 40, 128)	16512	[‘activation1’]
batch_normalization1_1	(None, 40, 128)	512	['conv1d1_1 [0] [0]']
Activation_1 (Activation)	(None, 40, 128)	0	['batch_normalization1_1]
conv1d1_2	(None, 40, 128)	49280	['activation1_1 [0] [0]']
Add (Add)	(None, 40, 128)	0	['conv1d_2[0][0]', ‘conv1d [0] [0]']
batch_normalization_2 (Batch Normalization)	(None, 40, 128)	512	[‘add [0] [0]']
Activation_2 (Activation)	(None, 40, 128)	0	['batch_normalization_2 [0] [0]']
conv1d_3 (Conv1D)	(None, 40, 128)	16512	['activation_2 [0] [0]']
batch_normalization_3	(None, 40, 128)	512	['conv1d_3 [0] [0]']
activation_3 (Activation)	(None, 40, 128)	0	['batch_normalization_3]
conv1d_4 (Conv1D)	(None, 40, 128)	49280	['activation_3 [0] [0]']
add_1 (Add)	(None, 40, 128)	0	['conv1d_4 [0][0]', ‘Add [0] [0]']
Max1_pooling1d (Max1Pooling1D)	(None, 13, 128)	0	['add_1 [0] [0]']
Dropout1_1 (Dropout1)	(None, 13, 128)	0	['max1_pooling [0] [0]']
Flatten1 (Flatten1)	(None, 1664)	0	['dropout_1 [0][0]']
Dense1_1 (Dense1)	(None, 5)	8325	[‘flatten [0][0]']

In summary, the methodology integrates genomic preprocessing, deep learning model construction (BiLSTM and ResNet), and performance evaluation. The following section presents the results of these models, comparing their predictive performance and analyzing their diagnostic potential.

4. Results and Discussion

The dataset comprises 60,000 protein sequences split into 70 percent, 10 percent, and 20 percent for training, validating, and testing purposes. The models are tested with a GPU processor and 16 Bytes of RAM. The model counts the number of codes, or amino acids, present in each sequence; typically, this ranges from 40 to 250 characters. Fig. 5 [28] displays the sequence counts.

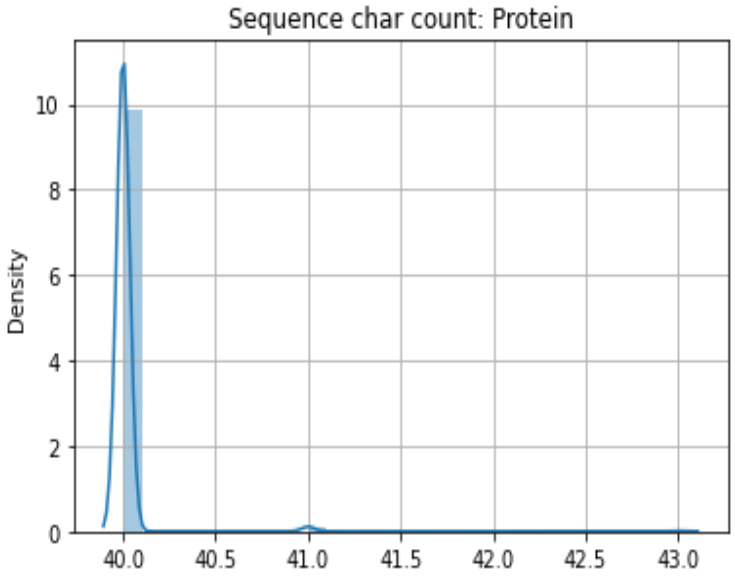

Fig.5: Sequence count Protein

Next, the model determines the frequency of each unaligned sequence's code; the findings are displayed in Fig. 6 [29, 30]. Training and Validation Accuracy, as well as Losses of the Bidirectional LSTM for the provided data, are shown in Fig. 7. Here are the training and validation accuracy and loss statistics for the provided data of ResNet CNN: Fig. 8.

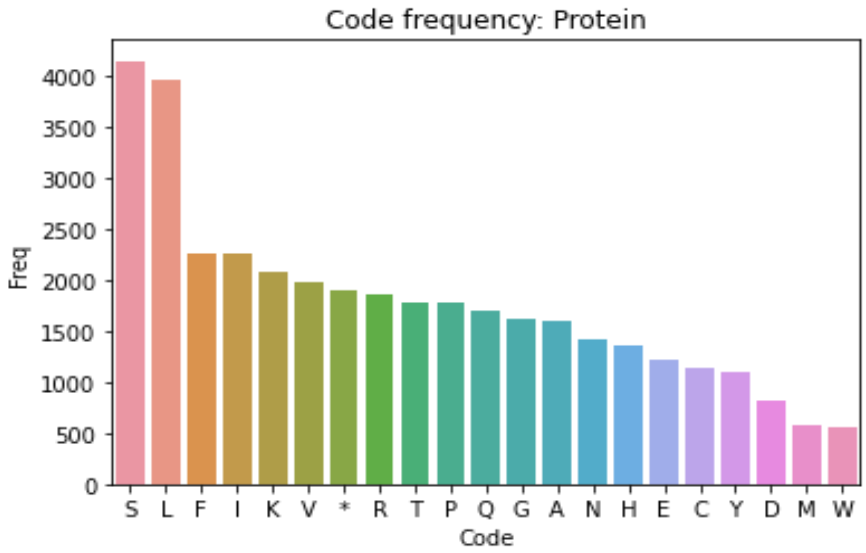

Fig.6: Training Code frequency protein

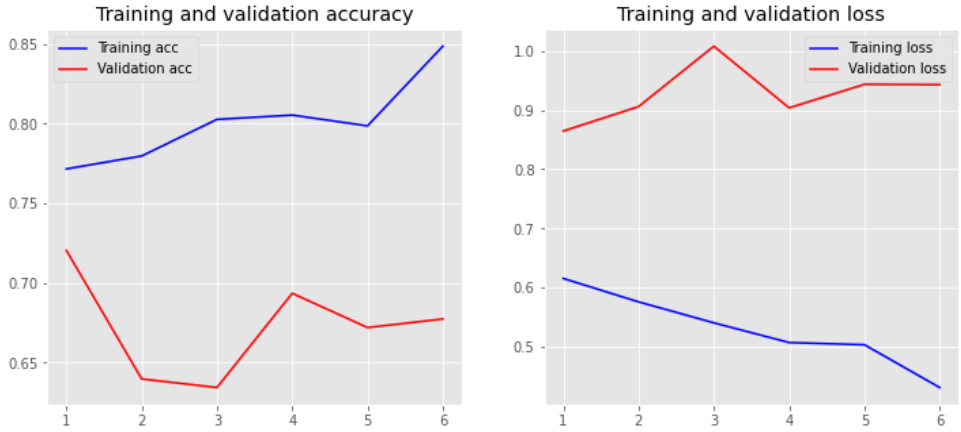

Fig.7: Bidirectional LSTM validation, training accuracy, and loss

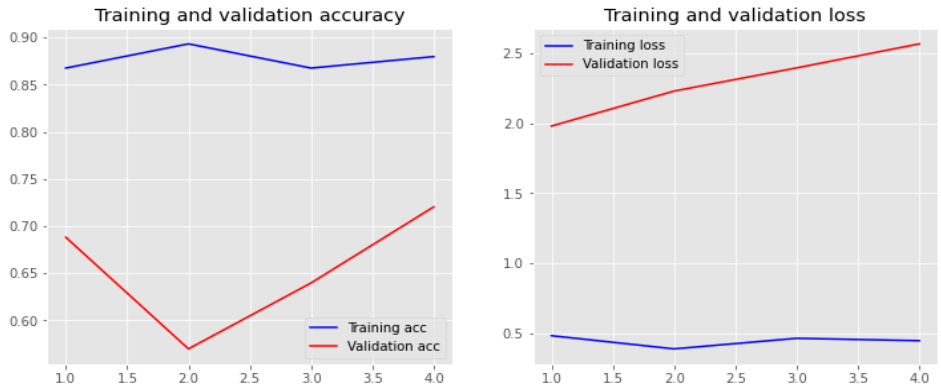

Fig.8: ResNet-based CNN training accuracy, validation, and loss

The data is classified using two different deep learning models. The model comparison is based on different parameters like accuracy, loss, val_accuracy, and val_loss. These are evaluated based on the number of epochs [31]. Table 4 shows the comparison matrix of these models.

Table 4. Proposed Model Comparison

Model	Epochs	Loss	Accuracy	Val_Loss	Val_accuracy
Bidirectional LSTM	10	0.6827	0.9068	0.9134	0.6828
	20	0.4836	0.7959	1.1380	0.6613
	30	0.8319	0.6919	0.8144	0.7527
ResNet-Based CNN	10	0.6449	0.7851	1.3600	0.7151
	20	0.5530	0.8797	1.9816	0.7204
	30	0.2866	0.9324	3.1818	0.6935

Ethical considerations are critical, particularly regarding bias in genomic data, which may disproportionately represent certain populations. Future datasets should ensure broader representation to avoid skewed diagnostic outcomes. Another concern is scalability—deep learning models require significant computational infrastructure, which may limit adoption in low-resource settings. Cloud-based deployment or

lightweight model compression (e.g., pruning, quantization) could improve accessibility. Finally, collaboration with clinical institutions is necessary to validate and benchmark these models against established diagnostic methods in real-world scenarios.

5. Conclusion

This paper introduced deep learning models for predicting Pleuro-pulmonary Blastoma by providing input data in protein sequences. These models extract features automatically from the protein sequences and provide a more accurate and efficient method for detecting the variant of PPB. This methodology enhances the connection between functional annotations of amino acid sequences and PPB classification. This paper compares two different deep learning models based on their performance: Bidirectional LSTM and ResNet-Based CNN. ResNet-based CNN outperforms Bidirectional LSTM in terms of both training and validation accuracy. Using these advanced algorithms leads to more precise diagnostic and prognostic capabilities and improves patient outcomes. In the future, this work can be extended by exploring hybrid deep learning models such as CNN-Transformers, which may capture more complex genomic patterns. Using multi-modal data that combines genomics with imaging (CT scans or histopathology) could further improve diagnostic accuracy. Expanding the dataset to include larger and more diverse patient groups will also help reduce bias and increase generalizability. Finally, close collaboration with clinical institutions will be essential to validate the models in real-world settings and support their integration into practical diagnostic workflows.

Acknowledgement

I would like to thank my colleagues for their support in this research work.

Conflict of Interest

The authors declare that there is no conflict of Interest.

References

[1] Bauer, A. J., Stewart, D. R., Kamihara, J., Harris, A. K., Turner, J., Shah, R., Schneider, K. W., Schneider, K., Carr, A. G., Harney, L. A., Frazier, A. L., Orbach, D., Schneider, D. T., Malkin, D., Dehner, L. P., Messinger, Y. H., Hill, D. A., & Schultz, K. A. P, DICER1 and Associated Conditions: Identification of At-risk Individuals and Recommended Surveillance Strategies-Response. Clinical cancer research : an official journal of the American Association for Cancer Research, 25(5), (2019) 1689–1690. https://doi.org/10.1158/1078-0432.CCR-18-3495

[2] Stewart, D. R., Best, A. F., Williams, G. M., Harney, L. A., Carr, A. G., Harris, A. K., Kratz, C. P., Dehner, L. P., Messinger, Y. H., Rosenberg, P. S., Hill, D. A., & Schultz, K. A. P, Neoplasm Risk Among Individuals With a Pathogenic Germline Variant in DICER1. Journal of clinical oncology : official journal of the American Society of Clinical Oncology, 37(8), (2019) 668–676. https://doi.org/10.1200/JCO.2018.78.4678.

[3] Brenneman, M., Field, A., Yang, J., Williams, G., Doros, L., Rossi, C., Schultz, K. A., Rosenberg, A., Ivanovich, J., Turner, J., Gordish-Dressman, H., Stewart, D., Yu, W., Harris, A., Schoettler, P., Goodfellow, P., Dehner, L., Messinger, Y., & Hill, D. A, Temporal order of RNase IIIb and loss-of-function mutations during development determines phenotype in pleuropulmonary blastoma / DICER1 syndrome: a unique variant of the two-hit tumor suppression model. F1000Research, 4, (2015) 214. https://doi.org/10.12688/f1000research.6746.2

[4] Messinger, Y. H., Stewart, D. R., Priest, J. R., Williams, G. M., Harris, A. K., Schultz, K. A., Yang, J., Doros, L., Rosenberg, P. S., Hill, D. A., & Dehner, L. P, Pleuropulmonary blastoma: a report on 350 central pathology-confirmed pleuropulmonary blastoma cases by the International Pleuropulmonary Blastoma Registry. Cancer, 121(2), (2015) 276–285. https://doi.org/10.1002/cncr.29032

[5] Hill, D. A., Jarzembowski, J. A., Priest, J. R., Williams, G., Schoettler, P., & Dehner, L. P, Type I pleuropulmonary blastoma: pathology and biology study of 51 cases from the international pleuropulmonary blastoma registry. The American journal of surgical pathology, 32(2), (2008) 282–295. https://doi.org/10.1097/PAS.0b013e3181484165

[6] López-Andreu, J. A., Ferrís-Tortajada, J., & Gómez, J, Pleuropulmonary blastoma and congenital cystic malformations. The Journal of pediatrics, 129(5), (1996) 773–775. https://doi.org/10.1016/s0022-3476(96)70176-2

[7] Sun, Y., Zhu, S., Ma, K., Liu, W., Yue, Y., Hu, G., Lu, H., & Chen, W, Identification of 12 cancer types through genome deep learning. Scientific reports, 9(1), (2019) 17256. https://doi.org/10.1038/s41598-019-53989-3

[8] Song, Y, CT Radio Genomics of Non-Small Cell Lung Cancer Using Machine and Deep Learning. 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), (2019) 128-139. https://doi.org/10.1109/ICCECE51280.2021.9342170.

[9] Adetiba, E., & Olugbara, O. O, Lung cancer prediction using neural network ensemble with histogram of oriented gradient genomic features. The-ScientificWorldJournal, 2015, (2015) 786013. https://doi.org/10.1155/2015/786013

[10] Zhang, B., Qi, S., Monkam, P., Li, C., Yang, F., Yao, Y., & Qian, W, Ensemble learners of multiple deep CNNs for pulmonary nodules Classification using CT images. IEEE Access, 7, (2019) 110358–110371. https://doi.org/10.1109/access.2019.2933670.

[11] Shaikh, F., & Rao, D, Prediction of Cancer Disease using Machine learning Approach. Materials Today Proceedings, 50, (2022) 40–47. https://doi.org/10.1016/j.matpr.2021.03.625

[12] Gunasekaran, H., Ramalakshmi, K., Rex Macedo Arokiaraj, A., Deepa Kanmani, S., Venkatesan, C., & Suresh Gnana Dhas, C, Analysis of DNA Sequence Classification Using CNN and Hybrid Models. Computational and mathematical methods in medicine, 2021, (2021) 1835056. https://doi.org/10.1155/2021/1835056

[13] Hu, H., Li, Z., Elofsson, A., & Xie, S, A Bi-LSTM based ensemble algorithm for prediction of protein secondary structure. Applied Sciences, 9(17), (2019) 3538. https://doi.org/10.3390/app9173538

[14] Sharma, B. S., Prabhakaran, V., Desai, A. P., Bajpai, J., Verma, R. J., & Swain, P. K, Post-translational Modifications (PTMs), from a Cancer Perspective: An Overview. Oncogen, 2(3), (2019). https://doi.org/10.35702/onc.10012

[15] Kao, H., Nguyen, V., Huang, K., Chang, W., & Lee, T, SuccSite: Incorporating Amino Acid Composition and Informative k-spaced Amino Acid Pairs to Identify Protein Succinylation Sites. Genomics Proteomics & Bioinformatics, 18(2), (2020) 208–219. https://doi.org/10.1016/j.gpb.2018.10.010

[16] Momenzadeh, M., Sehhati, M., & Rabbani, H, Using hidden Markov model to predict recurrence of breast cancer based on sequential patterns in gene expression profiles. Journal of Biomedical Informatics, 111, (2020) 103570. https://doi.org/10.1016/j.jbi.2020.103570

[17] Thapa, N., Chaudhari, M., McManus, S., Roy, K., Newman, R. H., Saigo, H., & Kc, D. B, DeepSuccinylSite: a deep learning based approach for protein succinylation site prediction. BMC Bioinformatics, 21(S3), (2020). https://doi.org/10.1186/s12859-020-3342-z

[18] Pandey, A., & Roy, S. S, Protein sequence classification using convolutional neural network and natural language processing. In Studies in big data, (2022) 133–144. https://doi.org/10.1007/978-981-16-9158-4_9

[19] C. Sekhar, K. Pavani and M. S. Rao, "Comparative analysis on Intrusion Detection system through ML and DL Techniques: Survey," 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA), Nagpur, India, 2021, pp. 1-5, doi: 10.1109/ICCICA52458.2021.9697291.

[20] Karagöz, M. A., & Nalbantoglu, O. U, Taxonomic classification of metagenomic sequences from Relative Abundance Index profiles using deep learning. Biomedical Signal Processing and Control, 67, (2021) 102539. https://doi.org/10.1016/j.bspc.2021.102539
[21] Deorowicz S, FQSqueezer: k-mer-based compression of sequencing data. Scientific reports, 10(1), (2020) 578. https://doi.org/10.1038/s41598-020-57452-6
[22] Roy, S. S., Sikaria, R., & Susan, A, A deep learning based CNN approach on MRI for Alzheimer's disease detection. Intelligent Decision Technologies, 13(4), (2020) 495–505. https://doi.org/10.3233/idt-190005
[23] Yu, L., Tanwar, D. K., Penha, E. D. S., Wolf, Y. I., Koonin, E. V., & Basu, M. K, Grammar of protein domain architectures. Proceedings of the National Academy of Sciences, 116(9), (2019) 3636–3645. https://doi.org/10.1073/pnas.1814684116
[24] Roy, S. S., & Taguchi, Y. H,. Identification of genes associated with altered gene expression and m6A profiles during hypoxia using tensor decomposition based unsupervised feature extraction. Scientific reports, 11(1), (2021) 8909. https://doi.org/10.1038/s41598-021-87779-7
[25] Balas, V. E., Roy, S. S., Sharma, D., & Samui, P, Handbook of Deep Learning Applications. Smart innovation, systems and technologies. (2019). https://doi.org/10.1007/978-3-030-11479-4
[26] A. M. Remita and A. B. Diallo, Statistical linear models in virus genomic alignment-free classification: application to hepatitis C viruses, in 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), San Diego, CA, USA, November 2019, (2019) 474-481, https://doi.org/10.1109/BIBM47256.2019.898337
[27] Zhang, D., & Kabuka, M. R, Protein Family Classification from Scratch: A CNN Based Deep Learning Approach. IEEE/ACM transactions on computational biology and bioinformatics, 18(5), (2021) 1996–2007. https://doi.org/10.1109/TCBB.2020.2966633
[28] Cheng, J., Liu, Y., & Ma, Y, Protein secondary structure prediction based on integration of CNN and LSTM model. Journal of Visual Communication and Image Representation, 71, (2020) 102844. https://doi.org/10.1016/j.jvcir.2020.102844
[29] Bileschi, M. L., Belanger, D., Bryant, D. H., Sanderson, T., Carter, B., Sculley, D., . . . Colwell, L. J, Using deep learning to annotate the protein universe. Nature Biotechnology, 40(6), (2022) 932–937. https://doi.org/10.1038/s41587-021-01179-w
[30] Sabapathy, D. G., Guillerman, R. P., Orth, R. C., Zhang, W., Messinger, Y., Foulkes, W., Priest, J. R., & Annapragada, A. V, Radiographic screening of infants and young children with genetic predisposition for rare malignancies: DICER1 mutations and pleuropulmonary blastoma. AJR. American journal of roentgenology, 204(4), (2015) W475–W482. https://doi.org/10.2214/AJR.14.12802
[31] Schultz, K. A., Harris, A., Williams, G. M., Baldinger, S., Doros, L., Valusek, P., Frazier, A. L., Dehner, L. P., Messinger, Y., & Hill, D. A, Judicious DICER1 testing and surveillance imaging facilitates early diagnosis and cure of pleuropulmonary blastoma. Pediatric blood & cancer, 61(9), (2014) 1695–1697. https://doi.org/10.1002/pbc.25092