Evaluating Performance of Machine Learning Models for Chromosomal Disorder Classification using Karyotype Images: A Comparative Study

Swathi Kamath
Department of MCA
NMAM Institute of Technology,
Nitte (Deemed to be) University,
Nitte, India
cswathikamath@gmail.com

Donald Ashwin Dsouza
Department of MCA
NMAM Institute of Technology,
Nitte (Deemed to be) University,
Nitte, India
donaldashwin@gmail.com

Prashanth Shetty D
K.S. Hegde Medical Academy,
Nitte (Deemed to be) University,
Mangalore, India
dprashanthshetty@gmail.com

Reshma Arun Shetty
K.S. Hegde Medical Academy,
Nitte (Deemed to be) University,
Mangalore, India
drreshma@nitte.edu.in

***Abstract*—This project presents a comprehensive study on the automated identification of chromosomal disorders using karyotype images. The goal is to enhance the accuracy and efficiency of diagnosis by leveraging machine learning techniques. The comparative analysis involves five prominent machine learning models, namely XGBoost, Random Forest, SVM, Naive Bayes, and InceptionV3. Additionally, two hybrid models are explored, combining InceptionV3 for feature extraction with either XGBoost or an Artificial Neural Network (ANN) for classification. The project tackles the challenge of limited publicly available datasets by collaborating with the Kshema Centre for Genetic Services, resulting in a dataset of 835 images. The results reveal that XGBoost and Random Forest models achieve high accuracy levels of 91% and 85%, respectively, outperforming other models. The hybrid models, combining InceptionV3 with XGBoost and ANN, also demonstrate promising performance, reaching accuracies of 78% and 60% respectively. These findings highlight the potential of hybrid models in capturing both the feature extraction capabilities of deep learning and the classification power of traditional machine learning algorithms. The successful implementation of these models paves the way for more accurate and efficient chromosomal disorder identification. By leveraging machine learning and hybrid modelling approaches, this project contributes to advancing medical image analysis and has significant implications for the field of chromosomal disorder diagnosis.**

***Index Terms*—Artificial Neural Network, Chromosomal Disorders, Feature Extraction, Hybrid Models, InceptionV3, Karyotype Images, Naive Bayes, Random Forest, SVM, XGBoost.**

I. Introduction

Chromosomes, the essential building blocks of human cell nuclei, which are essential for the transport of genetic information in the form of DNA. Humans typically contain 46 chromosomes, which are divided into 23 pairs. A single chromosome from each parent makes up each pair. Twenty-two of these pairs are autosomes, whereas the other pair consists of sex chromosomes. Males have one X and one Y chromosome (XY), while females normally have two X chromosomes (XX). Different genetic problems can result from variations or anomalies in the number or structure of these chromosomes.

To investigate and analyze chromosomal abnormalities, researchers and medical professionals employ a widely utilized technique known as karyotype analysis. This technique involves visual examination of an individual's chromosomes to identify any structural or numerical variations. The traditional methods employed for karyotype analysis heavily rely on human expertise, typically requiring trained cytogeneticists or medical professionals to manually inspect and interpret the results. Additionally, the process is time consuming, often taking several days to analyze a single sample. These limitations have necessitated the exploration of more efficient and accurate alternatives.

The field of medical image processing has taken quite a stride in the last couple of years since the approaches of deep learning and machine learning were synergized. The field of medical image processing has taken quite a stride in the last couple of years since the approaches of deep learning and machine learning were synergized. The purpose of this work is to offer a detailed comparative analysis of a few machine learning models in the light of the potential advantages offered to the field of karyotype analysis by machine learning. In this direction, the experiments aim to try out five most widely used models-two hybrid models, XGBoost, Random Forest, Support Vector Machines (SVM), Naive Bayes, and InceptionV3-on how accurately they can classify karyotype photos into four predefined categories: Down syndrome, Chronic Myeloid

Leukemia (CML), normal male, and normal female. With such evaluation of the models, this research aims to identify the best method to be used in the automation of classification of chromosomal disorders. Such advancements could revolutionize the field of karyotype analysis, enhancing efficiency, accuracy, and accessibility. Ultimately, this research has the potential to reduce dependence on manual methods, enabling timely interventions for individuals affected by chromosomal disorders.

II. Literature Survey

Ramaneswaran et al. [1] propose a hybrid model that combines Inception v3 and XGBoost for the classification of microscopic white blood cell images. In this model, XGBoost is utilized as the classification head, while Inception v3 serves as the feature extractor. The study demonstrates that replacing the softmax classification head with an XGBoost classification head improves the classification performance for several convolutional neural network (CNN) backbones. The hybrid model achieved a weighted F1 score of 0.986.

Bentéjac et al. [2] present a comparative analysis of XGBoost, random forests, and gradient boosting, focusing on training speed, generalization performance, and parameter tuning. The study concludes that XGBoost is a reliable and efficient machine learning algorithm, though it may not always be the optimal choice depending on the specific problem context.

Antonio et al. [3] provide an overview of hybrid artificial neural networks (ANNs) by categorizing them into three distinct perspectives: models, methods, and data. The paper emphasizes that ANNs, as part of computational intelligence, aim to model nonlinear systems resembling biological brain functions. The study explores various strategies for hybridizing neural networks, either through the combination of learning paradigms with gradient-based algorithms or through the integration of heterogeneous data structures.

Al-Kharraz et al. [4] focus on the identification of metaphase images using the YOLO v2 object detection model, which separates individual chromosomes. Feature extraction and classification are performed using transfer learning with the VGG19 model, which is fine-tuned by incorporating global average pooling (GAP) layers, fully connected (FC) layers, and batch normalization (BN) layers. The reported accuracy for chromosome identification on the CEGMR dataset is 95.04%, with a final detection accuracy of 96.67% for identifying numerical abnormalities in classes 13, 18, 21, and X.

Moradi and Setarehdan [5] utilized a dataset comprising 303 chromosomal images from classes 16, 17, and 18, sourced from the Cytogenetic Laboratory of the Cancer Institute at Imam Hospital in Tehran, Iran. The authors constructed a six-dimensional feature vector by combining the width and position of the two most visually significant areas of each chromosome, along with its length and centromeric index. Using artificial neural networks, they achieved an accuracy of 98.6% in classifying the three chromosomal classes.

Poletti et al. [6] proposed a feature extraction method that involves calculating the area, perimeter, length, and contour profile of chromosomes, as well as measuring the density and contour of 64 samples. These features were then fed into an ANN for classification. To improve classification, the authors applied a greedy strategy to reorganize and reassign the 24 chromosomal classes. The method was tested on a publicly available dataset of 5,474 chromosomal images from the Biomedical Imaging Laboratory (BioImLab), yielding an average accuracy of 94%.

Swati et al. [7] integrated convolutional super-resolution and Xception networks into a unified system for chromosome image classification. The authors applied a convolutional super-resolution network as a preprocessing step to enhance low-resolution images by mapping them to high-resolution counterparts after length normalization. Subsequently, the high-resolution images were inputted into a CNN classification system based on the Xception architecture. The proposed method achieved a classification accuracy of 92.36% on the BioImLab dataset.

Sharma and Vig [8] utilized Residual Neural Networks (ResNet-50) for feature extraction from convolutional layers. These features were then fed into a Long Short-Term Memory (LSTM) network enhanced with an attention block. A fully connected softmax layer was employed for classifying chromosomes into one of 24 categories. The approach was evaluated on a publicly available dataset from the Biomedical Imaging Laboratory, consisting of 5,474 chromosome images. The method achieved a classification accuracy of 90.42%.

Kusakci et al. [9] employed the Copenhagen dataset, which includes data extracted from 4,400 chromosome images. To reduce the dimensionality of the feature set, the authors applied Principal Component Analysis (PCA). For classification, they trained a separate Support Vector Machine (SVM) for each pair of chromosomes. The method demonstrated an impressive classification accuracy of 97.84% across 22 chromosome types.

Roshtkhari and Setarehdan [10] focused on classifying chromosomes in group E, specifically chromosomes 16, 17, and 18, using manually designed features. These features included the centromeric index, relative length, and characteristics derived from the chromosome's density profile, which were extracted using discrete wavelet transform. Dimensionality reduction of the feature space was performed via Linear Discriminant Analysis (LDA). Classification was carried out using a three-layer feed-forward perceptron neural network, resulting in an average classification accuracy of 99.3%.

Zhang et al. [11] acquired an image set of 224 karyotyping samples from a local company and employed a region filtering technique to remove noise without preprocessing the dataset. Using the regionprop function in Matlab, bounding boxes were drawn around each segmented chromosome. A CNN architecture, consisting of five layers (convolution, pooling, dropout, flatten, and fully connected layers), was designed for feature extraction and classification. The method achieved an accuracy of 92.5% on the PWCK dataset. The network was

trained on vertical chromosome orientations, with future work planned to address various chromosome orientations.

Mashadi and Seyedin [12] focused on chromosomal normalization techniques, specifically length normalization and intensity normalization, rather than directly extracting features from the images. After applying these normalization techniques, the authors trained an SVM classifier using pixel-level image data as input. The dataset consisted of 42,000 manually separated chromosome images obtained from the Royan Institute. The method achieved a classification accuracy of 95.9% when length normalization was applied.

Vanitha and Venmathi [13] extracted both textural and global features from a dataset containing 4,600 manually segmented chromosome images. The textural features were derived from the Gray Level Co-occurrence Matrix (GLCM), including contrast, inverse difference moment, angular second moment, correlation, variance, and homogeneity. Global features included relative length, relative area, and centromeric index. The system consisted of two phases. In the first phase, a Self-Organizing Map (SOM) Neural Network was employed to classify chromosomes into seven categories (A-G). In the second phase, based on the initial classification, a hybrid neural network combining K-Means, Learning Vector Quantization (LVQ), and Naïve Bayes classifiers with serial fusion was applied to classify the chromosomes into 24 categories. The system achieved a classification accuracy of 98%, and the authors noted its ability to detect most anomalies in metaphase samples, although the specific technique for anomaly identification was not discussed.

Gagula-Palalic and Can [14] proposed an ensemble approach known as Competitive Neural Network Teams (CNNTs), which integrates a set of 462 simple perceptrons arranged in 22 x 21 learning machines. These perceptrons were trained to distinguish between two classes. The authors used the Sarajevo dataset, which includes chromosomal features for 3,300 chromosomes collected from the Clinical Center of the University of Sarajevo. The dataset comprises 10 primary features generated from band pattern vectors, including the length of the short p-arm of the chromosomes. The CNNT approach demonstrated an error rate of 1.73% in classifying the chromosomes into 22 distinct categories.

Somasundaram [15] utilized a dataset consisting of 1,000 touching chromosomes, 1,000 overlapping chromosomes, and 500 multiple overlapping chromosomes, with both normal and aberrant examples. Two segmentation techniques were employed, including the MOGAC approach, which was specifically used to isolate overlapping chromosomes, although it could not distinguish between overlapping and touching chromosomes. The study did not evaluate the performance of the segmentation process. Chromosomes were classified into 24 categories using Probabilistic Neural Networks (PNN) and Support Vector Machines (SVM). The classifier used chromosomal features such as length, centromere index, and similarity index. The SVM classifier achieved an accuracy of 97%, while the PNN classifier achieved 96% accuracy, with PNN outperforming SVM in this context.

Rungruangbaiyok and Phukpattaranont [16] preprocessed chromosome images by removing noise and enhancing contrast and image quality. They segmented 60 metaphase images using thresholding and Otsu's method. Additional image processing techniques, such as dilation and erosion, were applied to improve the images, although the effectiveness of these segmentation techniques was not assessed. Feature extraction was performed using methods such as singular value decomposition, perimeter, band profile, and band area. The classification system was based on a Probabilistic Neural Network (PNN), which was divided into two phases: the first phase classified the chromosomes into six groups based on extracted features, and the second phase grouped the six categories into 24 classes. The classification achieved an accuracy of 61.30% for males and 68.18% for females.

Sharma et al. [17] used a dataset of chromosome images collected from 400 metaphase samples of healthy patients. The authors employed non-expert crowd-sourcing via CrowdFlower to assist in chromosome separation. Several preprocessing techniques were applied, including bending orientation and center determination, chromosome straightening, chromosome sewing, reconstruction, and length normalization. The proposed CNN architecture for feature extraction and classification consisted of four stacked blocks of two convolutional layers, one dropout layer, and one max-pooling layer, followed by a 24-unit softmax layer and two fully connected layers. Without preprocessing, the CNN achieved an accuracy of 68.5%, which increased to 86.7% when preprocessing techniques were applied.

Markou et al. [18] implemented a four-stage process to generate feature vectors for chromosome classification. In the first stage, each pixel along the medial axis of the chromosome was scored. In the second stage, line segments were drawn for each pixel, extending perpendicular to the derivative vector at that point. The grayscale intensity values of the pixels below each line segment were then calculated. These statistics were used to derive the median value for each segment. The dataset used for the experiment consisted of 4,554 expertly segmented chromosomal images of healthy individuals from the Papageorgiou Regional General Hospital's Molecular Biology Laboratory in Thessaloniki, Greece. For classification into 24 categories, the authors used a hybrid two-level classification approach, with the first phase involving context-independent SVM classification and the second phase applying a context-aware post-classification stage. The overall error rate for this method was 6.65%.

III. Methodology

A. Data Acquisition

This project utilizes a dataset of karyotype images collected from the Kshema Centre for Genetic Services, Derlakatte. Due to the unavailability of publicly accessible datasets for karyotype analysis, a collaboration was established with the Kshema Centre for Genetic Services to procure the required images for the study. The collaboration involved visiting the laboratory and collecting the images by referring to the

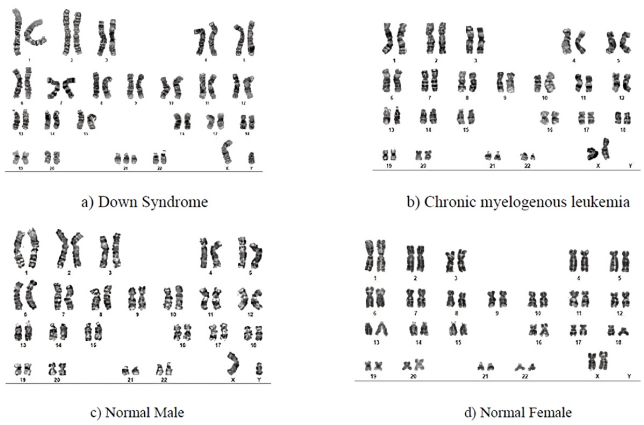

Fig. 1. Sample Data

medical records. The dataset comprises karyotype images from diverse individuals, encompassing four distinct classes: normal male, normal female, Chronic Myeloid Leukemia (CML), and Down syndrome. The four classes represent common chromosomal conditions encountered in clinical practice. Since the dataset was collected specifically for this project, efforts were made to ensure the inclusion of a representative sample of karyotype images for each class. However, the dataset could be subject to certain limitations, such as variations in the number of images per class or potential bias due to the specific medical cases available for collection. The dataset acquisition process involved adhering to ethical considerations and privacy regulations to protect the identity and confidentiality of the individuals associated with the medical records. The dataset was obtained with proper consent and following the necessary protocols. It is important to note that for this project, the knowledge base was built from scratch, as there were no existing datasets readily applicable to the task of automated chromosomal disorder identification. The acquisition of this unique dataset enabled the development and evaluation of machine learning models specifically designed for the task at hand.

B. Preprocessing

In order to prepare the karyotype pictures for analysis and classification, preprocessing is essential. The subsequent preprocessing methods are utilized to improve the caliber and appropriateness of the pictures: Grayscale Conversion: The original color format of the karyotype photos is converted to grayscale through the process of grayscale conversion. By reducing each image to a single channel and doing away with the complications that come with several color channels, this conversion makes the analysis that follows easier to understand. 2. Resize: A standard size of 299 by 299 pixels is applied to the photos. By maintaining consistency in image dimensions, resizing eliminates bias resulting from differences in image sizes and enables consistent processing. 3. Reshaping: The purpose of reshaping the arrays is to flatten the image features into a 2-dimensional format, where each row represents an image and each column represents a feature. This is often done to prepare the data for input to a machine learning model that expects a 2-dimensional feature matrix.

C. Training

Four traditional machine learning models, namely SVM, Random Forest, Naive Bayes, and XGBoost, were trained using the extracted features from the pre-processed data. These models were trained using their respective algorithms and hyperparameter settings on the training set. The feature vectors obtained from the preprocessing step were used as inputs for these models, which learned to classify the karyotype images into their respective classes based on the provided training labels.

In addition to the traditional models, an InceptionV3 model, a deep learning architecture, was employed for training. The InceptionV3 model was trained directly on the pre-processed images, bypassing the feature extraction step. It underwent a training process involving 20 epochs, during which the model learned to extract hierarchical features from the input images and make predictions based on these features.

Additionally, two hybrid models that combined the benefits of deep learning and conventional machine learning techniques were designed. For the first hybrid model, feature extraction was performed using the InceptionV3 model. Subsequently, the InceptionV3 model was used for feeding the preprocessed images with the derived features being sent to the XGBoost model for further training. This hybrid model utilized the excellent classification abilities of the XGBoost model along with the excellent feature extraction ability of the InceptionV3 model.Like the first hybrid model that utilized a different approach, the second hybrid model used the InceptionV3 model to extract the features. Again, after preprocessing the photos, they were inserted into the InceptionV3 model to carry out the feature extraction. Then an ANN was trained on the feature extracted.This hybrid model harnessed the feature extraction capabilities of InceptionV3 and the flexibility and learning capabilities of the ANN to achieve accurate classification of the karyotype images.

The models' internal parameters were iteratively changed during training in order to reduce the loss function and enhance prediction accuracy. The model parameters were updated using optimization procedures, including gradient descent, in accordance with the training data. Accuracy, precision, recall, and F1 score were among the performance metrics that the models were trained to maximize on the training set.

Evaluation and Comparison of Algorithms

Performance indicators are crucial for assessing the karyotype analysis system's accuracy and efficacy.When evaluating the effectiveness of categorization models, the following measures are frequently employed:

1) Accuracy: The overall accuracy of the system's predictions is measured by accuracy. It is calculated by dividing the total number of correct predictions by the total number of predictions made. But accuracy by

itself might not give a whole picture of performance, particularly when there are unequal class distributions.

2) Precision:The percentage of true positive predictions among all positive forecasts is quantified by precision.It indicates the model's ability to avoid false positives. The ratio of true positives to the total of true positives and false positives is used to compute precision.
3) Recall (Sensitivity): The percentage of true positives that the model properly detected is measured by recall, which is also referred to as sensitivity."It reflects the model's ability to minimize false negatives. The ratio of true positives to the total of true positives and false negatives is used to compute recall.
4) F1 Score: The harmonic mean of recall and precision is the F1 score.It provides a balanced measure of a model's performance by considering both recall and precision. The F1 score runs from 0 to 1, with 1 signifying the best possible performance. It is computed as two times the precision and recall product divided by the sum of the two.

MODELS USED

1) Support Vector Machine (SVM): An SVM is a high-dimensional feature space hyperplane-building non-linear model for classification. It attempts to create an efficient division along with maximizing the margin between different groups. To transform features of the input into a higher-dimensional space where classes can be linearly separated, SVMs use kernel functions. The SVM picks a boundary decision which best separates the classes during training. The model then classifies the new samples by the position of the sample relative to the decision boundary into which class they belong.

2) Random Forest: An ensemble learning model, called Random Forest, is an aggregate collection of decision trees used to make predictions. Each decision tree is fitted on a random subset of the training data and a random subset of the characteristics. For the chosen features, each of the trees learns a decision rule independently at training. Voting or averaging the predictions made by each individual tree gives the final output in prediction. Randomness helps encourage generalization and prevent overfitting in feature selection and in data sampling.

3) Naive Bayes: Naive Bayes is the probabilistic classification model derived from Bayes' theorem. The word "naive" describes the assumption that the features are conditionally independent given the class. Naive Bayes derives the probabilities of each class and makes its forecasts based on the feature values using the one with the highest probability. The model works based on calculations of class probabilities using the training data and then applies them to newly discovered instances for classification. Naive Bayes has good performance in high-dimensional feature spaces and is computationally efficient.

TABLE I
SUMMARY OF MODELS USED

Models Used	Methodology	Future Scope
SVM	Constructs a hyperplane in high-dimensional space, utilizing kernel functions for non-linear separations.	Designing kernels and optimization algorithms more efficiently to improve multi-class classification performance.
Random Forest	Utilizes an ensemble of decision trees trained on random subsets of data and features. Outputs based on majority voting or averaging.	Implementing advanced methods of ensemble techniques to improve the interpretability and computational efficiency for handling large datasets.
Naive Bayes	Probabilistic model based on Bayes' theorem with a "naive" assumption of feature independence.	Introducing dependencies between features for better classification accuracy and performance in complex datasets.
XGBoost	Gradient boosting framework that iteratively adds weak learners (decision trees) to reduce loss. Regularization is added for better generalization.	Expanding to advanced boosting frameworks like LightGBM and CatBoost to reduce training time and improve efficiency.
InceptionV3	A deep CNN, trained from inception on large datasets like ImageNet on a hierarchical feature-extraction basis to learn, with fine-tuning conducted on the particular classification task.	Applying transfer learning to similar datasets and fine-tuning to enhance robustness in specific classification tasks.
XceptionBoost	Hybrid model combining InceptionV3 for feature extraction and XGBoost for classification, leveraging strengths of both.	Enhancing feature fusion techniques to reduce training time without compromising on accuracy and robustness.
IncepANN	Uses InceptionV3 for feature extraction, followed by an ANN for flexible classification. Optimized for non-linear decision boundaries.	Implementing regularization strategies and testing with diverse data to improve generalizability and robustness.

4) XGBoost: XGBoost is an algorithm of gradient boosting: highly performing ensemble learning using the combinations of multiple weak prediction models, classic being decision trees.It enables the iterative addition of new trees to the ensemble during the training process hence allowing the faults made by the preceding trees to be repaired. XGBoost uses the gradient descent optimization technique to optimize a given loss function. Regularization techniques have been incorporated in XGBoost to reduce overfitting and improve generalization. The last prediction is obtained by combining the individual predictions of the number of trees at large in the ensemble.

5) InceptionV3 (Convolutional Neural Network): A deep convolutional neural network architecture called InceptionV3 has been pre-trained using a sizable image dataset called ImageNet. Convolutional and pooling procedures are layered

several times, and then fully connected layers are added. The goal of InceptionV3 is to capture intricate and hierarchical visual patterns in pictures. The network gains the ability to identify and encode pertinent information from the input images during training. Fine-tuning is done to adapt the pre-trained InceptionV3 model to the specific objective of chromosomal abnormality classification. The anticipated class probabilities are produced by the model's last layer.

6) XceptionBoost (Hybrid Model): This is a novel approach for classification of karyotypes. In the first part of the architecture, InceptionV3 is used for feature extraction.The ImageNet dataset served as the pre-training set for the deep convolutional neural network (CNN) model InceptionV3. The model leverages InceptionV3's capacity to automatically identify and extract pertinent characteristics from the input photos. The feature vectors obtained from the InceptionV3 feature extraction process are fed into the XGBoost classifier as input. With its ability to handle structured data with precision and efficiency, XGBoost is a powerful gradient boosting method.It can handle numerous input features and is especially good at handling complicated categorization problems. The XGBoost classifier takes the retrieved features as input and performs the classification operation. It employs boosting techniques and an ensemble of decision trees to iteratively enhance the model's performance. XGBoost can handle classification issues with several classes and provide robust and accurate predictions.

By integrating InceptionV3 for feature extraction and XGBoost for classification, the hybrid model takes advantage of both the deep learning capabilities of CNNs and the boosting power of XGBoost. This architecture allows for the effective utilization of extracted features to make accurate predictions and achieve improved performance in the classification of karyotype images. Overall, the hybrid model architecture brings together the strengths of both InceptionV3 and XGBoost to create a robust and accurate system for chromosomal disorder classification.

7) IncepANN (Hybrid Model): The architecture of this hybrid model involves two main components: InceptionV3 for feature extraction and a Sequential model of Artificial Neural Network (ANN) for classification.

InceptionV3 is a CNN which retrieves good information from the input photographs. Then it forwards the features to the feature representation stage that generates a one-dimensional feature vector. The data is now simplified in this step so it can be processed further.The data is passed to the Artificial Neural Network consisting of many interconnected layers of nodes in the end. Using the learned features, the pattern and prediction ability of ANN is used for classification. The process of training refines the model's accuracy and performance by iteratively optimizing the weights and bias within ANN. The hybrid model can be applied to inference and prediction after it has been trained. InceptionV3 processes new images to extract features, which are subsequently sent to the trained ANN for classification. Predicted class labels or a probability distribution across the classes are the model's outputs. All things considered, the hybrid model creates an architecture that is able to accurately and efficiently classify images by fusing the feature extraction powers of InceptionV3 with the pattern identification and classification powers of the ANN.

IV. Result

Figure 2 shows the performance of various machine learning models in classifying chromosomal disorders based on karyotype images. The accuracy of each model provides an indication of how well it correctly identifies and categorizes different chromosomal disorders.

XGBoost achieved the highest accuracy of 91%, indicating its strong ability to accurately classify the given karyotype images. This model outperformed the others and is considered the most effective for this specific task.

SVM achieved an accuracy of 77%, demonstrating its ability to classify chromosomal disorders with reasonable accuracy. Although lower than XGBoost, SVM still performed well in the classification task.

Random Forest achieved an accuracy of 85%, indicating its effectiveness in accurately identifying chromosomal disorders. This ensemble model showed promising results and demonstrated good performance in the classification task.

Naive Bayes achieved an accuracy of 63%, which is relatively lower compared to the other models. This suggests that Naive Bayes may be less effective at capturing the complexities and patterns in the karyotype images for accurate classification.

InceptionV3, a convolutional neural network model, achieved an accuracy of 62%. While it performed adequately, its accuracy was lower than that of the other models. This indicates that InceptionV3 may not have been able to effectively learn and extract the relevant features from the karyotype images for accurate classification.

The hybrid models, XceptionBoost and IncepANN, achieved accuracies of 78% and 60%, respectively. These models combine the feature extraction capabilities of InceptionV3 with other algorithms such as XGBoost and artificial neural networks. While they showed promising results, their accuracies were slightly lower compared to XGBoost.

V. Conclusion

In conclusion, this project represents a significant step toward the automated identification of chromosomal disorders from karyotype images using machine learning models. The development of this system was motivated by the scarcity of publicly available datasets and the time-consuming nature of traditional diagnostic methods. By collaborating with the Kshema Centre for Genetic Services and collecting a comprehensive dataset, these challenges were effectively addressed, resulting in the creation of a robust knowledge base from scratch.

Through the comparative analysis of various machine learning models, including XGBoost, SVM, Random Forest, Naive Bayes, and InceptionV3, the potential of these models for

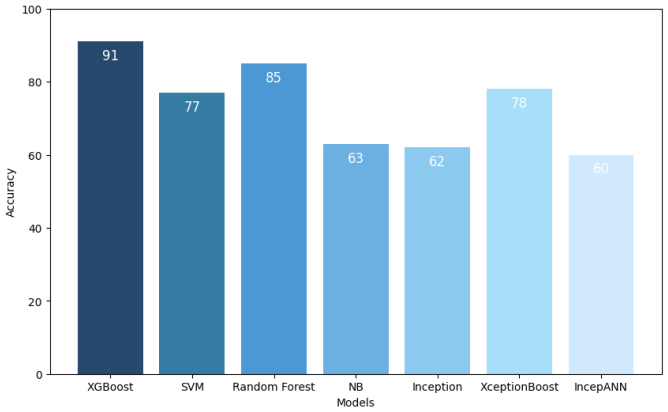

Fig. 2. Accuracy comparison of machine learning models

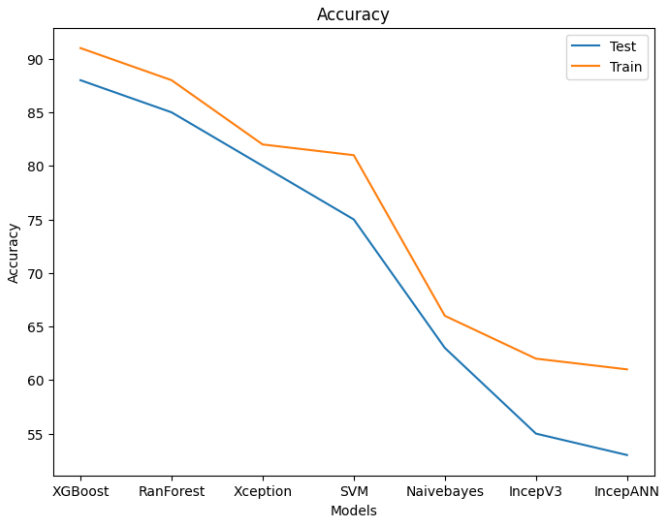

Fig. 3. Train vs. Test accuracy across models

accurately classifying chromosomal disorders has been explored. As shown in Table II, the results revealed that XGBoost and Random Forest emerged as the top-performing models, demonstrating superior performance in accurately classifying chromosomal disorders. These models surpassed traditional methods and even deep learning approaches, achieving impressive accuracies of 91% and 85%, respectively. This underscores the effectiveness of ensemble-based approaches and decision tree algorithms in addressing the complexities associated with chromosomal disorder classification.

Although the hybrid models, utilizing InceptionV3 for feature extraction and subsequent classification via XGBoost and ANN, yielded promising results, further optimization and fine-tuning are required to unlock their full potential. Furthermore, while the current dataset offers valuable insights, the availability of larger and more diverse datasets could significantly enhance the robustness and generalization capabilities of these models. The integration of more varied and extensive data would likely lead to models that can better generalize across different populations and types of chromosomal abnormalities.

TABLE II
PERFORMANCE METRICS FOR VARIOUS MODELS

Model	Precision	Recall	F1-Score	Support
XGBOOST				
CML	0.86	0.82	0.84	38
Down Syndrome	0.97	0.95	0.96	38
Normal Female	1.00	0.94	0.97	44
Normal Male	0.93	0.87	0.90	47
Accuracy			0.91	
SVM				
CML	0.68	0.74	0.71	38
Down Syndrome	0.75	0.79	0.77	38
Normal Female	0.91	0.93	0.92	44
Normal Male	0.85	0.70	0.77	47
Accuracy			0.77	
Random Forest				
CML	0.76	0.66	0.70	38
Down Syndrome	0.97	0.89	0.93	38
Normal Female	0.81	0.98	0.89	44
Normal Male	0.87	0.85	0.86	47
Accuracy			0.85	
Naïve Bayes				
CML	0.50	0.68	0.58	38
Down Syndrome	0.68	0.58	0.63	38
Normal Female	0.78	0.64	0.70	44
Normal Male	0.69	0.53	0.60	47
Accuracy			0.63	
Inception-V3				
CML	0.75	0.79	0.77	42
Down Syndrome	0.78	0.87	0.83	40
Normal Female	0.51	0.47	0.49	43
Normal Male	0.62	0.57	0.59	44
Accuracy			0.62	
Xception Boost				
CML	0.73	0.79	0.76	38
Down Syndrome	0.78	0.76	0.77	38
Normal Female	0.84	0.82	0.83	44
Normal Male	0.78	0.77	0.78	47
Accuracy			0.78	
Incep-ANN				
CML	0.73	0.71	0.72	38
Down Syndrome	0.78	0.76	0.77	38
Normal Female	0.64	0.52	0.57	44
Normal Male	0.56	0.51	0.53	47
Accuracy			0.60	

One of the key limitations encountered in this study is the relatively small size and scope of the dataset, which may have

constrained the generalization of the models. Additionally, the complexity of chromosomal disorders and the nuances in karyotype images present challenges in achieving consistently high accuracy across all cases. Further efforts are needed to incorporate a broader range of chromosomal abnormalities and image types to improve the models' adaptability to real-world clinical settings.

Nevertheless, this effort lays a solid foundation for future developments in medical image processing and diagnostics. By combining automated technologies with clinical workflows, there is significant potential to improve patient outcomes, reduce human error, and increase the overall effectiveness of genetic condition detection. The integration of machine learning models with existing clinical practices can facilitate faster, more accurate diagnoses, ultimately leading to early interventions and better patient care.

REFERENCES

[1] Ramaneswaran, S. & Srinivasan, Kathiravan & Vincent P M, Durai & Chang, Chuan-Yu. (2021). Hybrid Inception v3 XGBoost Model for Acute Lymphoblastic Leukemia Classification. Computational and Mathematical Methods in Medicine. 2021. 1-10. 10.1155/2021/2577375.

[2] Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2018). A Comparative Analysis of XGBoost. arXiv preprint arXiv:1803.05743.

[3] Gutiérrez, Pedro Antonio & Martínez, Cesar. (2011). Hybrid Artificial Neural Networks: Models, Algorithms and Data. 6692. 177-184. 10.1007/978-3-642-21498-1_23.

[4] M. S. Al-Kharraz, L. A. Elrefaei and M. A. Fadel, "Automated System for Chromosome Karyotyping to Recognize the Most Common Numerical Abnormalities Using Deep Learning," in IEEE Access, vol. 8, pp. 157727-157747, 2020

[5] M. Moradi and S. K. Setarehdan, "New features for automatic classification of human chromosomes: A feasibility study", Pattern Recognit. Lett., vol. 27, no. 1, pp. 19-28, Jan. 2006.

[6] E. Poletti, E. Grisan and A. Ruggeri, "A modular framework for the automatic classification of chromosomes in Q-band images", Comput. Methods Programs Biomed., vol. 105, no. 2, pp. 120-130, Feb. 2012.

[7] S. Swati, M. Sharma and L. Vig, "Automatic classification of low-resolution chromosomal images", Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops, pp. 315-325, 2019.

[8] M. Sharma, Swati, and L. Vig, "Automatic chromosome classification using deep attention based sequence learning of chromosome bands," in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2018, pp. 1–8.

[9] A. O. Kusakci, B. Ayvaz, and E. Karakaya, "Towards an autonomous human chromosome classification system using competitive support vector machines teams (CSVMT)," Expert Syst. Appl., vol. 86, pp. 224–234, Nov. 2017.

[10] M. J. Roshtkhari and S. K. Setarehdan, "Linear discriminant analysis of the wavelet domain features for automatic classification of human chromosomes," in Proc. 9th Int. Conf. Signal Process., Oct. 2008, pp. 849–852.

[11] W. Zhang, S. Song, T. Bai, Y. Zhao, F. Ma, J. Su, and L. Yu, "Chromosome classification with convolutional neural network based deep learning," in Proc. 11th Int. Congr. Image Signal Process., Biomed. Eng. Informat.

[12] N. Tabatabaey Mashadi and S. Alireza Seyedin, "Direct classification of human G-banded chromosome images using support vector machines," in Proc. 9th Int. Symp. Signal Process. Appl., Feb. 2007, pp. 1–4.

[13] L. Vanitha and A. R. Venmathi, "Automatic abnormality detection in chromosomes using hybrid multi-layer neural network," Int. J. Neural Netw. Appl., vol. 4, pp. 37–45, 2011.

[14] S. Gagula-Palalic and M. Can, "Human chromosome classification using competitive neural network teams (CNNT) and nearest neighbor," in Proc. IEEE-EMBS Int. Conf. Biomed. Health Informat. (BHI), Jun. 2014, pp. 626–629.

[15] D. Somasundaram, "Structural similarity and probabilistic neural network based human G-band chromosomes classification," Int. J. Hum. Genet., vol. 18, no. 3, pp. 228–237, Apr. 2018.

[16] S. Rungruangbaiyok and P. Phukpattaranont, "Chromosome image classification using a two-step probabilistic neural network," Songklanakarin J. Sci. Technol., vol. 32, no. 3, pp. 255–262, 2010.

[17] M. Sharma, O. Saha, A. Sriraman, R. Hebbalaguppe, L. Vig, and S. Karande, "Crowdsourcing for chromosome segmentation and deep classification," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Jul. 2017, pp. 34–41

[18] C. Markou, C. Maramis, A. Delopoulos, C. Daiou, and A. Lambropoulos, "Automatic chromosome classification using support vector machines," in Pattern Recognition: Methods and Applications. Hong Kong: iConceptPress, 2012, pp. 1–24.