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

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.

Introduction

Chromosomes, the fundamental structures found within the nucleus of human cells, play a pivotal role in carrying genetic information in the form of DNA. In the case of humans, the typical number of chromosomes is 46, arranged into 23 pairs. Each pair consists of one chromosome inherited from each parent. Among these pairs, 22 are autosomes, while the remaining pair comprises sex chromosomes. Females typically possess two X chromosomes (XX), while males have one X and one Y chromosome (XY). Deviations or abnormalities in the number or structure of these chromosomes can lead to various genetic disorders.

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.

In recent years, the field of medical image processing has witnessed remarkable progress through the integration of deep learning and machine learning techniques. In light of the potential advantages that machine learning offers in the domain of karyotype analysis, the objective of this project is to conduct a comprehensive comparative analysis of different machine learning models. Specifically, the study aims to evaluate the performance of five prominent models: XGBoost, Random Forest, Support Vector Machines (SVM), Naive Bayes, InceptionV3 and two hybrid models. The focus lies on their ability to accurately classify karyotype images into four distinct categories: normal male, normal female, Chronic Myeloid Leukemia (CML), and Down syndrome. By scrutinizing the performance of these models, the project seeks to identify the most effective approach for automating the 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.

Literature Survey

Ramaneswaran et al. [1] propose a hybrid Inception v3 XGBoost model for the classification of acute lymphoblastic leukemia (ALL) from microscopic white blood cell images. The proposed model uses Inception v3 as the image feature extractor and XGBoost as the classification head. The authors demonstrate that using an XGBoost classification head instead of a softmax classification head improves classification performance for this dataset for several different CNN backbones (feature extractors). The proposed hybrid model achieves a weighted F1 score of 0.986.

Bentéjac et al. [2] presents a comparative analysis of XGBoost, random forests, and gradient boosting. The analysis includes training speed, generalization performance, and parameter setup. The paper concludes that XGBoost is a reliable and efficient machine learning challenge solver, but it may not necessarily be the best choice under all circumstances.

Antonio et al. [3] provide a review of hybrid artificial neural networks (ANNs) from three perspectives: models, algorithms, and data. ANNs are flexible nonlinear models designed to mimic biological neural systems and are one of the three main components of computational intelligence. Hybrid ANNs have been proposed by extending the standard approach of ANNs, combining gradient-based algorithms with other learning paradigms, or using heterogeneous data structures. The paper presents different methods used for hybridizing neural networks, grouped using three criteria: models, algorithms, or data.

Al-Kharraz et al. [4] used YOLO v2 for detecting individual chromosomes on metaphase images. They used transfer learning with a pre-trained VGG19 model for feature extraction and classification, combining the Global Average Pooling (GAP) layer, Fully Connected (FC) layers, and Batch Normalization (BN) layer to fine-tune the VGG19. The approach achieved a 95.04% classification accuracy on the CEGMR dataset. The final numerical abnormality detection step achieved a 96.67% detection accuracy in the 13, 18, 21, and X classes.

Moradi and Setarehdan [5] obtained their dataset from the Cytogenetic Laboratory of the Cancer Institute at Imam Hospital in Tehran, Iran, comprising 303 bent chromosomal images from classes 16, 17, and 18. They created a six-dimensional feature by combining the width and position of the two most visually appealing areas of each chromosome with its length and centromeric index. ANN was used to achieve a 98.6% accuracy in classifying the three classes.

Poletti et al. [6] developed a feature vector based on the chromosome's area perimeter, length, area, and 64 samples for the density and contour profile. For classification, they fed these features into an ANN followed by a greedy approach to reassign and rearrange the 24 classes. They tested their approach on the publicly available Biomedical Imaging Laboratory (BioImLab) dataset of 5474 chromosome images, achieving an average accuracy of 94%.

Swati et al. [7] integrated convolutional super-resolution and Xception networks for their system. They performed length normalization as a preprocessing step and applied a convolutional super-resolution network to convert low-resolution images to high-resolution images. Xception CNN was then used for classification. The proposed system was evaluated on the BioImLab dataset, achieving a 92.36% classification accuracy.

In their study, Sharma and Vig [8] used Residual Neural Networks (ResNet-50) to extract features from convolutional layers. These features were then inputted into a Long Short-Term Memory (LSTM) network with an attention block. Finally, a fully connected softmax layer was used to classify chromosomes into one of 24 categories. The authors evaluated their approach on a publicly available dataset of chromosome images from the Biomedical Imaging Laboratory, which consisted of 5,474 images. They reported achieving a classification accuracy of 90.42%.

Kusakci et al. [9] conducted a study where they utilized the Copenhagen dataset consisting of features extracted from 4400 chromosomes. In order to reduce the number of input features, they applied Principal Component Analysis (PCA). They used Support Vector Machines (SVMs) to

train a single SVM for each pair of chromosomes. The proposed method achieved an accuracy of 97.84% in correctly classifying 22 chromosomes.
Roshtkhari and Setarehdan [10] utilized handcrafted features to classify chromosomes into (16, 17, and 18) classes in group E. The authors extracted features from the density profile of the chromosome using the discrete wavelet transform and also considered the centromeric index and relative length. To reduce the feature space, they applied Linear Discriminant Analysis. A three-layer feed-forward perceptron neural network was then used to classify the chromosomes, resulting in an average correct classification rate of 99.3%.

Zhang et al. [11] obtained the dataset, which included 224 karyotyping photos, from a nearby company and used the area filter to eliminate noise from the photographs before preprocessing them. They generated bounding boxes on each chromosome they isolated using regionprop in Matlab, and they did not assess segmentation performance. CNN was built with feature extraction and categorisation in mind. The five types of layers that made up the system were convolution, pooling, dropout, flatten, and dense layers. They obtained 91.3% of the well-classified karyotype (PWCK) with 92.5% accuracy. They used vertical chromosomes to train this network, and they will take into account chromosome orientations in the future.

Mashadi and Seyedin [12] employed two types of chromosomal normalization—intensity and length normalization—instead of extracting features from chromosome images. They then used the image pixels as the input pattern to the Support Vector Machine (SVM) classifier, which classified 24 classes. Their dataset, which they obtained from the Royan Institute, consists of 42,000 manually isolated chromosomal pictures. Through testing, they discovered that applying length normalisation produced a higher classification accuracy of 95.9%.

Vanitha and Venmathi [13] extracted textural features from GLCM (contrast, inverse difference moment, angular sec-ond moment, correlation, variance, and homogeneity) and global features (relative length, relative area, and centromeric index) from a dataset that included 4600 manually segmented chromosome images.There were two steps in the system. In the first step, a Self Organising Map Neural Network was used to classify chromosomes into seven groups (A-G). Using a hybrid neural network strategy that incorporates K-Mean, LVQ, and Naïve Bayes with a serial fusion, the seven groups grouped into 24 classes using the correctly identified chromosomes from the first step. They reported a 98% classification accuracy and the ability to identify several anomalies in the metaphase sample, but they didn't specify the precise method used to identify the abnormalities.

Gagula-Palalic and Can [14] suggested a Competitive Neu- ral Network Teams (CNNTs) that ensemble of ANN and near- est neighbor classifiers. There are 462 simple perceptrons in this approach. The 22 x 21 learning machines are perceptrons that have each been trained to discriminate between two classes.

They made use of the Sarajevo dataset, which was collected from the Clinical Centre of the University of Sarajevo and contained chromosomal characteristics for 3300 chromosomes. The dataset includes 10 main components derived from band pattern vectors, as well as the length of the short p-arm and chromosomes. With an error rate of 1.73%, they were able to classify chromosomes into 22 groups.

The dataset used by Somasundaram [15] included 1000 touching, 1000 overlapping, and 500 multiple overlapping chromosomes with both normal and aberrant examples. They preprocessed the metaphase image by morphologically operating (dilation, erosion, opening, and closure) and applying a median filter to lower the noise. Two segmentation techniques are used. The first technique, the MOGAC approach, was used to isolate overlapped chromosomes since it was unable to distinguish between overlapping and touching chromosomes. Moreover, this study did not assess segmentation performance. Chromosomes were classified into 24 classes using PNN and SVM. Chromosome length, centromere index, and similarity index are used as input features by the classifier in SVM classification, which yielded a 97% accuracy rate. On the other hand, they achieved 96% accuracy for PNN using chromosomal length, centromere position, and perimeter characteristics. The study revealed that PNN outperformed SVM.

Rungruangbaiyok and Phukpattaranont [16] began by preprocessing the images, which entails eliminating noise and enhancing contrast and quality of the image. With the use of Otsu's method and thresholding, they segmented sixty metaphase images. The photos were enhanced using dilation and erosion processing methods. The report doesn't address how to measure the segmentation process' performance. Following segmentation, they extracted features related to singular value decomposition, perimeter, band profile, and area of the band. Their system was divided into two phases, and they used the Probabilistic Neural Network (PNN). Based on traits that were extracted, the chromosomes were divided into six groups. Classifying the six categories into 24 groups was the next stage. For females, the accuracy result was 68.18%, and for males, it was 61.30%.

Sharma et al. [17] used non-expert crowding from CrowdFlower to isolate chromosomes from 400 healthy patient metaphases that made up their dataset, which they obtained from a hospital. Before classifying the chromosomes, they underwent a few preprocessing procedures such as bending orientation determination, bending centre determination, chromosome straightening, chromosome stitching, reconstruction, and chromosomal length normalisation. Four blocks make up the CNN they built for feature extraction and classification; each block has two convolutional layers, one dropout layer, and one maxpooling layer. Two fully connected layers and a 24-unit softmax layer come next. They achieved 68.5% classification accuracy without preprocessing steps and increased the accuracy to 86.7% with preprocessing procedures.

Markou et al. [18] constructed the feature vector by following four steps. They went through every pixel on the chromosome's medial axis in the first step. Second, they determined the line segment in each pixel that starts at that specific place and is orientated perpendicular to the deriva-tive vector at that location. Subsequently, computing the greyscale intensity values of the pixels situated beneath the line segment. At last, the median has been calculated from those numbers. The dataset used in this work comes from the General Regional Hospital Papageorgiou's Laboratory of Molecular Biology in Thessaloniki, Greece. It comprises 4554 expertly segmented chromosomal pictures of healthy individuals. To categorise 24 classes, they used a hybrid 2-level classification approach. A context-independent SVM classification process took place at the first level followed by a context-aware post-classification stage. They got an overall 6.65% classification errors.

Methodology

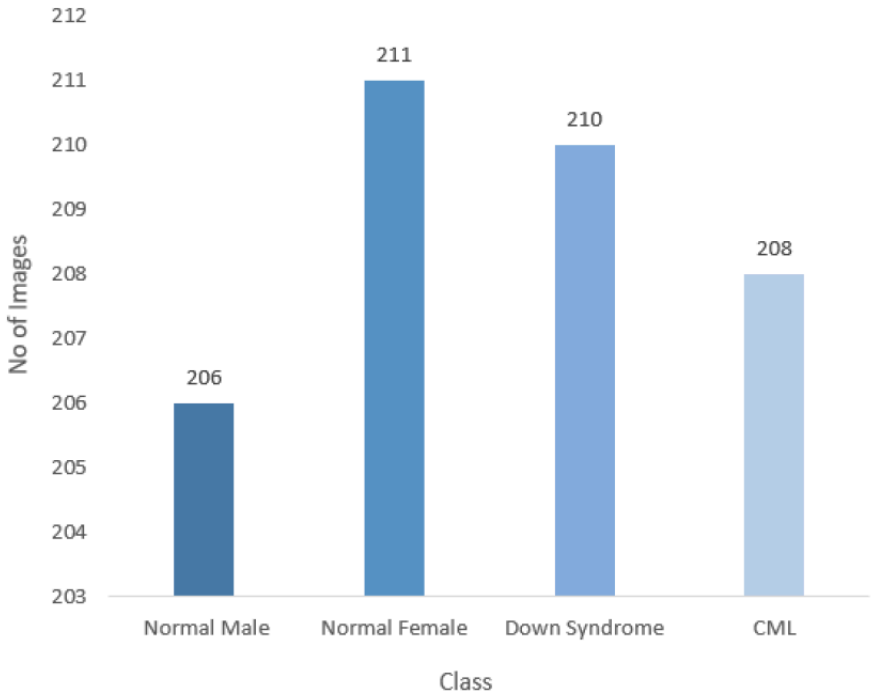
Data Acquisition

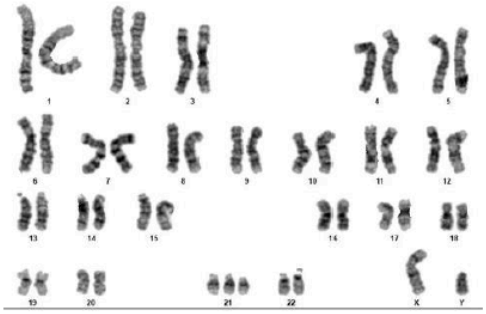
The dataset used in this project consists of karyotype images collected from the Kshema Centre for Genetic Services, Derlakatte. Due to the unavailability of publicly accessible datasets for karyotype analysis, a collaboration was established with the Kshema Centre for Genetic Services to procure the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

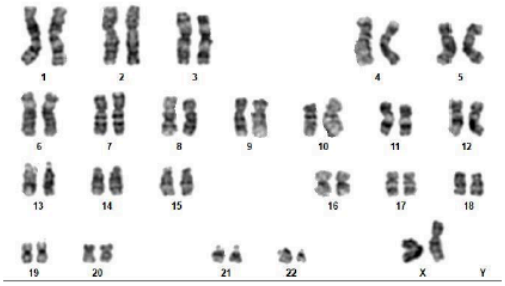
The dataset acquisition process involved adhering to ethical considerations and privacy regulations to protect the identity and confidentiality of the individuals associated with the medical records. The dataset was obtained with proper consent and following the necessary protocols.

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 allowed for the development and evaluation of machine learning models tailored specifically to the task at hand.

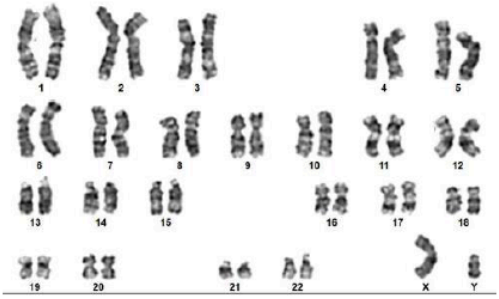




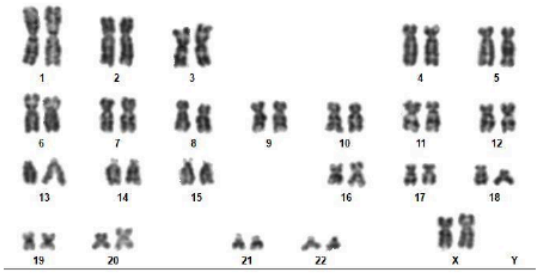
a) Down Syndrome



b) Chronic myelogenous leukemia



c) Normal Male



d) Normal Female

Preprocessing

Preprocessing plays a crucial role in preparing the karyotype images for analysis and classification. The following preprocessing techniques are employed to enhance the quality and suitability of the images:

1. Grayscale Conversion: The karyotype images undergo grayscale conversion, transforming them from their original colour format to grayscale. This conversion simplifies the subsequent analysis by reducing each image to a single channel, eliminating complexities associated with multiple colour channels.
2. Resize: The images are resized to a standardized size, typically 299x299 pixels. Resizing ensures uniformity in image dimensions, allowing for consistent processing and preventing any bias caused by variations in image sizes.
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.

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.

Two hybrid models were also developed by combining the strengths of traditional machine learning and deep learning approaches. The first hybrid model utilized the InceptionV3 model for feature extraction. The pre-processed images were fed into the InceptionV3 model, and the extracted features were then passed to the XGBoost model for further training. This hybrid model combined the superior feature extraction capabilities of the InceptionV3 model with the robust classification abilities of XGBoost.

The second hybrid model also employed the InceptionV3 model for feature extraction but used a different approach. The pre-processed images were again fed into the InceptionV3 model to extract features, and these features were then passed to an Artificial Neural Network (ANN) for

training. 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.

During the training process, the models adjusted their internal parameters iteratively to minimize the loss function and improve their predictive performance. Optimization algorithms, such as gradient descent, were employed to update the model parameters based on the training data. The models were trained to optimize performance metrics, including accuracy, precision, recall, and F1 score, on the training set.

Evaluation and Comparison of Algorithms

Performance metrics are essential for evaluating the effectiveness and accuracy of the karyotype analysis system. The following metrics are commonly used to assess the performance of classification models:

1. Accuracy: Accuracy measures the overall correctness of the system's predictions. It is calculated as the ratio of the number of correct predictions to the total number of predictions. However, accuracy alone may not provide a complete picture of performance, especially in the presence of imbalanced class distributions.

2. Precision: Precision quantifies the proportion of true positive predictions among all positive predictions. It represents the model's ability to avoid false positives. Precision is calculated as the ratio of true positives to the sum of true positives and false positives.

3. Recall (Sensitivity): Recall, also known as sensitivity, measures the proportion of true positives correctly identified by the model. It represents the model's ability to avoid false negatives. Recall is calculated as the ratio of true positives to the sum of true positives and false negatives.

4. F1 Score: The F1 score is the harmonic mean of precision and recall. It provides a balanced measure of a model's performance by considering both precision and recall. The F1 score ranges from 0 to 1, with 1 indicating the best possible performance. It is calculated as 2 times the product of precision and recall divided by the sum of precision and recall.

Performance Metric	Equation
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
Accuracy	$\frac{TP + TN}{TP + FN + TN + FP}$
F1-score	$\frac{2.(precision.recall)}{precision + recall}$

Models used

Support Vector Machine (SVM)

SVM is a non-linear classification model that constructs a hyperplane in a high-dimensional feature space. It aims to maximize the margin between different classes, allowing for effective separation. SVMs utilize kernel functions to transform the input features into a higher-dimensional space, where the classes can be linearly separable. During training, the SVM learns a decision boundary that optimally separates the classes. During prediction, the model assigns new samples to the appropriate class based on their position relative to the decision boundary.

Random Forest

Random Forest is an ensemble learning model that combines multiple decision trees to make predictions. Each decision tree is trained on a random subset of the training data and a random subset of the features. During training, each tree independently learns its own decision rules based on the selected features. During prediction, the final output is obtained through voting or averaging the predictions of the individual trees. The randomness in feature selection and data sampling helps reduce overfitting and improve generalization.

Naive Bayes

Naive Bayes is a probabilistic classification model based on Bayes' theorem. It assumes that the features are conditionally independent given the class, which is known as the "naive" assumption. Naive Bayes calculates the probabilities of each class based on the feature values and selects the class with the highest probability as the prediction. The model estimates the class probabilities using the training data and applies them to new instances for classification. Naive Bayes is computationally efficient and works well with high-dimensional feature spaces.

XGBoost

XGBoost is a gradient boosting algorithm that combines multiple weak prediction models, typically decision trees, to create a powerful ensemble model. During training, XGBoost iteratively adds new trees to the ensemble, with each tree correcting the errors of the previous ones. It uses a gradient descent optimization technique to minimize a specified loss function. XGBoost incorporates regularization techniques to prevent overfitting and achieve better generalization. The final prediction is obtained by aggregating the predictions of all the trees in the ensemble.

InceptionV3 (Convolutional Neural Network)

InceptionV3 is a deep convolutional neural network architecture that has been pre-trained on a large-scale image dataset, such as ImageNet. It consists of multiple layers of convolutional and pooling operations, followed by fully connected layers. InceptionV3 is designed to capture hierarchical and complex visual patterns in images. During training, the network learns to extract and encode relevant features from the input images. Fine-tuning is applied to adapt the pre-trained InceptionV3 model to the specific task of chromosomal disorder classification. The final layer of the model outputs the predicted class probabilities.

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. InceptionV3 is a deep convolutional neural network (CNN) model that is pre-trained on the ImageNet dataset. It is capable of extracting high-level features from images, which are essential for understanding the content and patterns present in the data. By utilizing InceptionV3, we leverage its ability to automatically learn and extract relevant features from the input images.

After extracting the features using InceptionV3, the feature vectors are passed as input to the XGBoost classifier. XGBoost is a powerful gradient boosting algorithm known for its accuracy and efficiency in handling structured data. It is particularly effective in handling complex classification tasks and can handle a large number of input features. The XGBoost classifier takes the extracted features as input and performs the classification task. It utilizes an ensemble of decision trees and applies boosting techniques to iteratively improve the model's performance.

XGBoost is capable of handling multi-class classification problems and provides robust and accurate predictions.

By combining 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.

IncepANN (Hybrid Model)

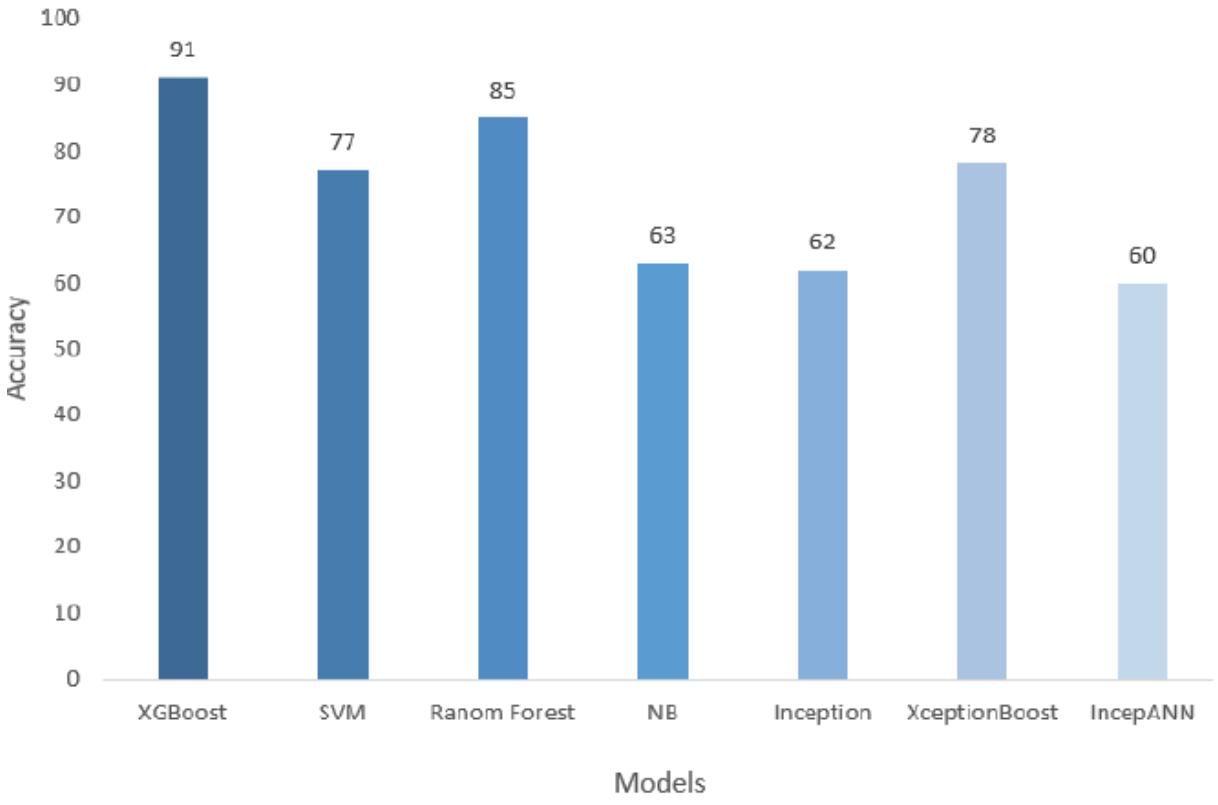
The architecture of this hybrid model involves two main components: InceptionV3 for feature extraction and a Sequential model of Artificial Neural Network (ANN) for classification.

InceptionV3, a pre-trained convolutional neural network (CNN), is used to extract meaningful features from input images. It captures various visual patterns, textures, and structures present in the images, providing a rich representation of the input data. The extracted features are then passed through a feature representation step, which transforms them into a one-dimensional feature vector. This step simplifies the data and prepares it for further processing. The feature vector is then fed into the ANN, which consists of multiple interconnected layers of nodes. The ANN performs classification tasks based on the extracted features, leveraging its ability to recognize patterns and make predictions. During training, the weights and biases of the ANN are adjusted through an iterative optimization process, enhancing the model's accuracy.

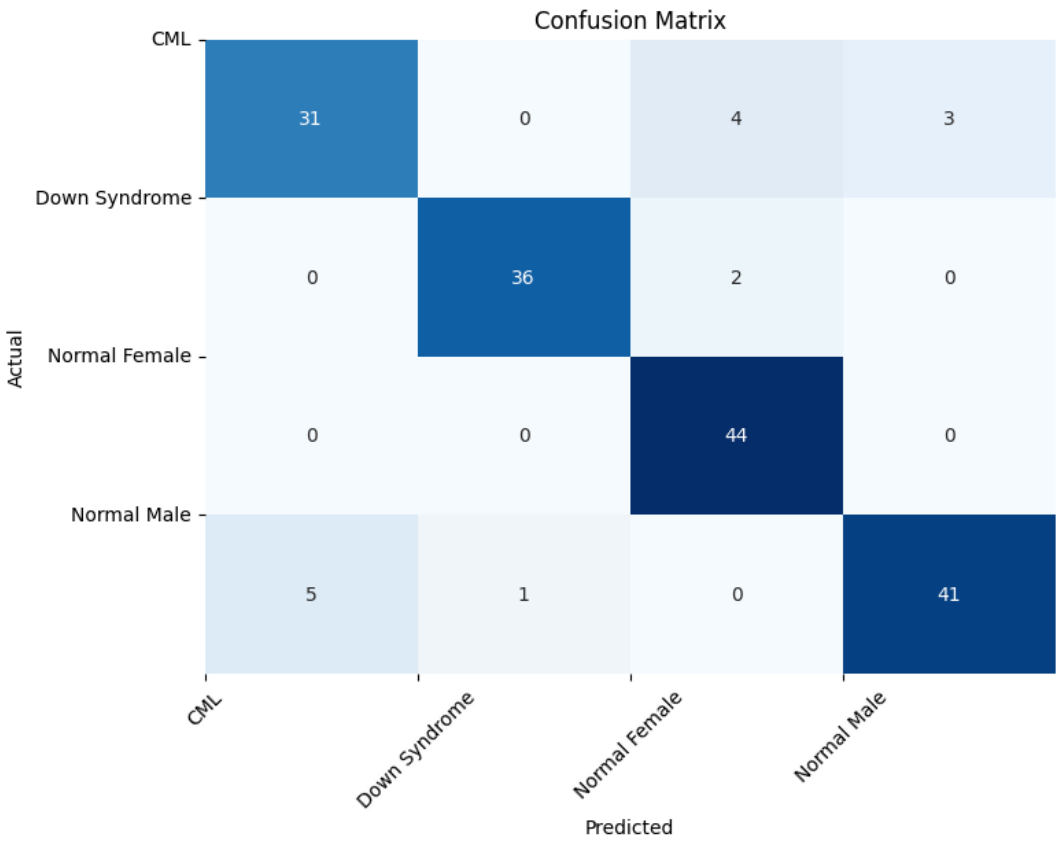
Once the hybrid model is trained, it can be used for inference and prediction. New images are processed by InceptionV3 to extract features, and these features are then passed to the trained ANN for classification. The output of the model is the predicted class label or a probability distribution over the classes. Overall, the hybrid model combines the feature extraction capabilities of InceptionV3 with the pattern recognition and classification abilities of the ANN, resulting in an architecture that is capable of accurate and efficient image classification.

Results

The results of this project show 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 can be considered the most effective in this specific task.

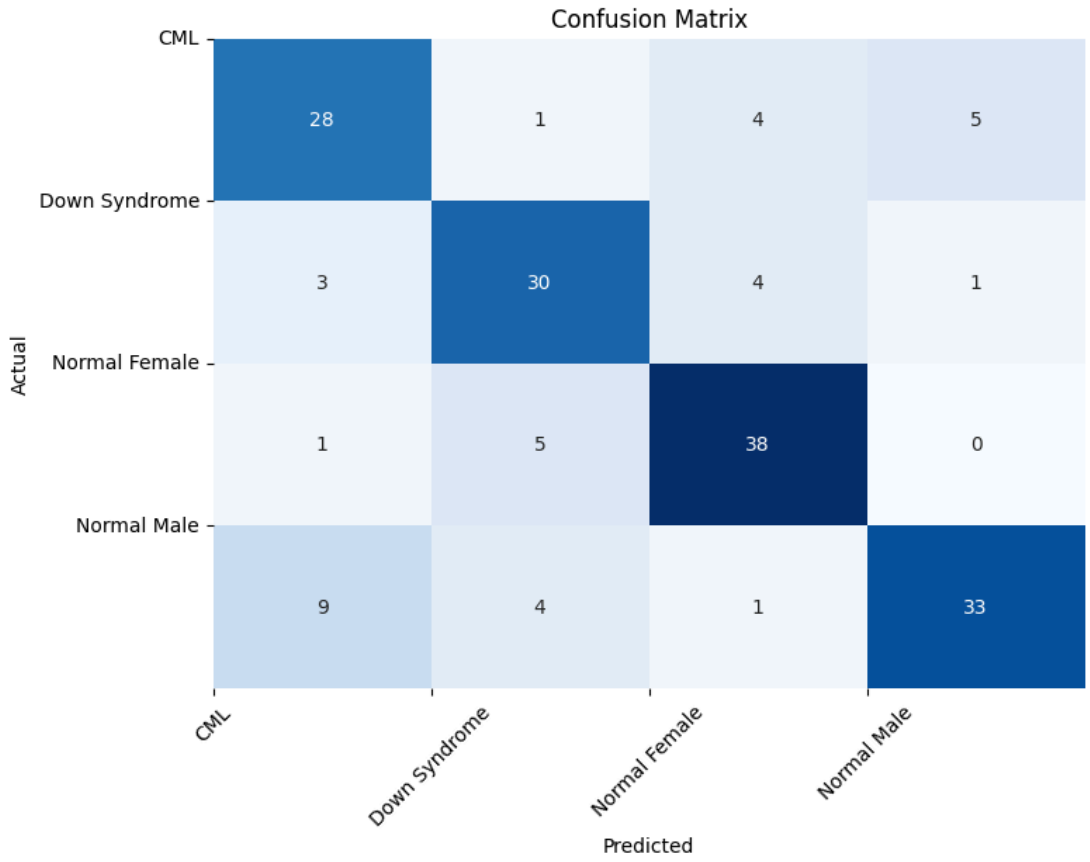


Confusion matrix for XGBoost

	precision	recall	f1-score	support
CML	0.86	0.82	0.84	38
Down Syndrome	0.97	0.95	0.96	38
Normal Female	0.88	1.00	0.94	44
Normal Male	0.93	0.87	0.90	47
accuracy			0.91	167
macro avg	0.91	0.91	0.91	167
weighted avg	0.91	0.91	0.91	167

Performance metrices for XGBoost

SVM achieved an accuracy of 77%, demonstrating its capability to classify chromosomal disorders with a reasonable level of accuracy. Although not as high as XGBoost, SVM still performed well in the classification task.

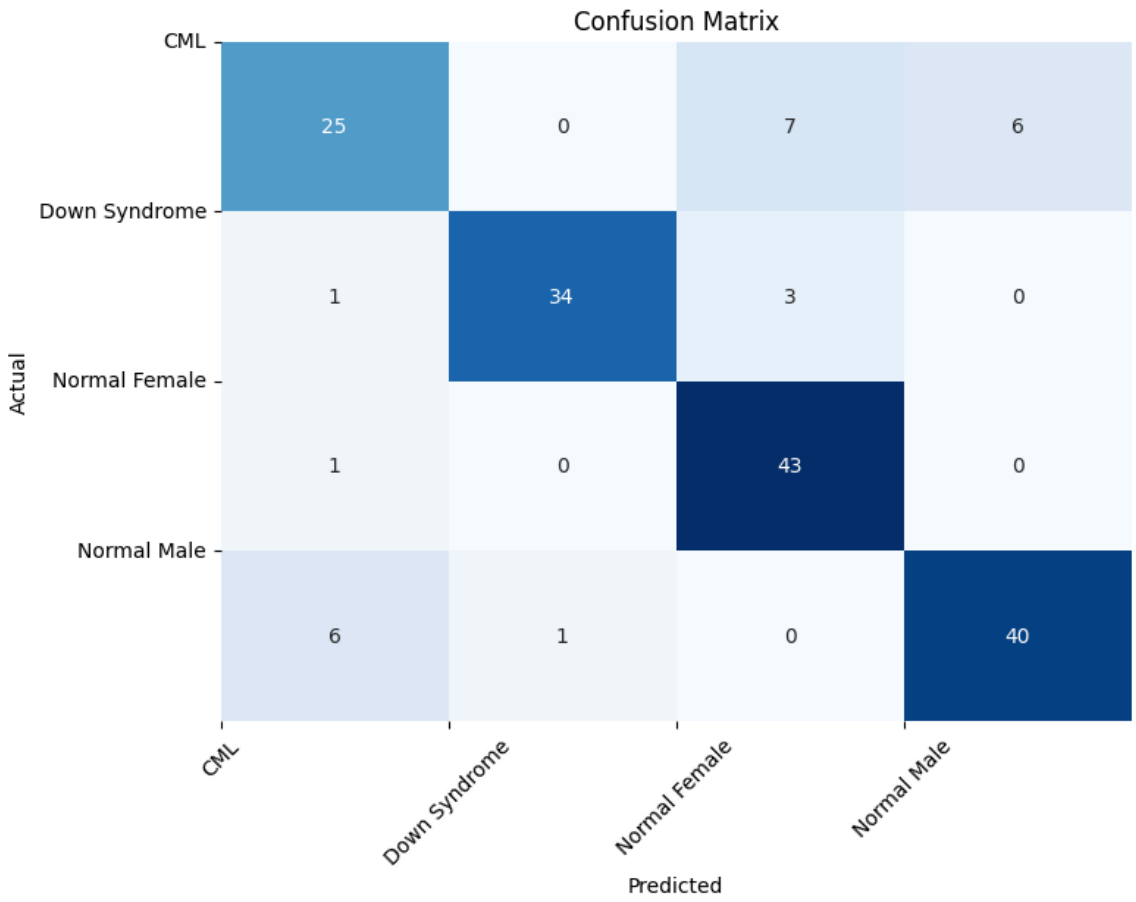


Confusion matrix for SVM

	precision	recall	f1-score	support
CML	0.68	0.74	0.71	38
Down Syndrome	0.75	0.79	0.77	38
Normal Female	0.81	0.86	0.84	44
Normal Male	0.85	0.70	0.77	47
accuracy			0.77	167
macro avg	0.77	0.77	0.77	167
weighted avg	0.78	0.77	0.77	167

Performance metrices for SVM

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.

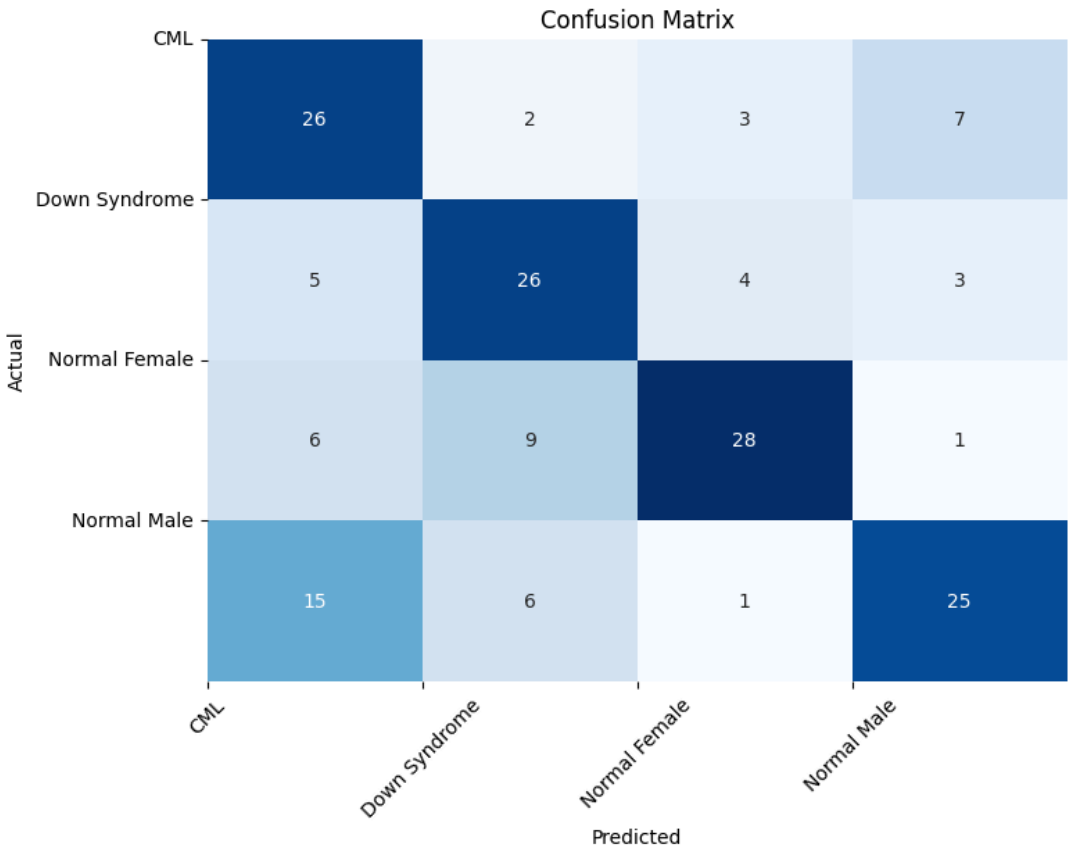


Confusion matrix for Random Forest

	precision	recall	f1-score	support
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	167
macro avg	0.85	0.85	0.85	167
weighted avg	0.85	0.85	0.85	167

Performance metrices for Random Forest

Naive Bayes achieved an accuracy of 63%, which is relatively lower compared to the other models. This suggests that Naive Bayes may not be as effective in capturing the complexities and patterns present in the karyotype images for accurate classification.

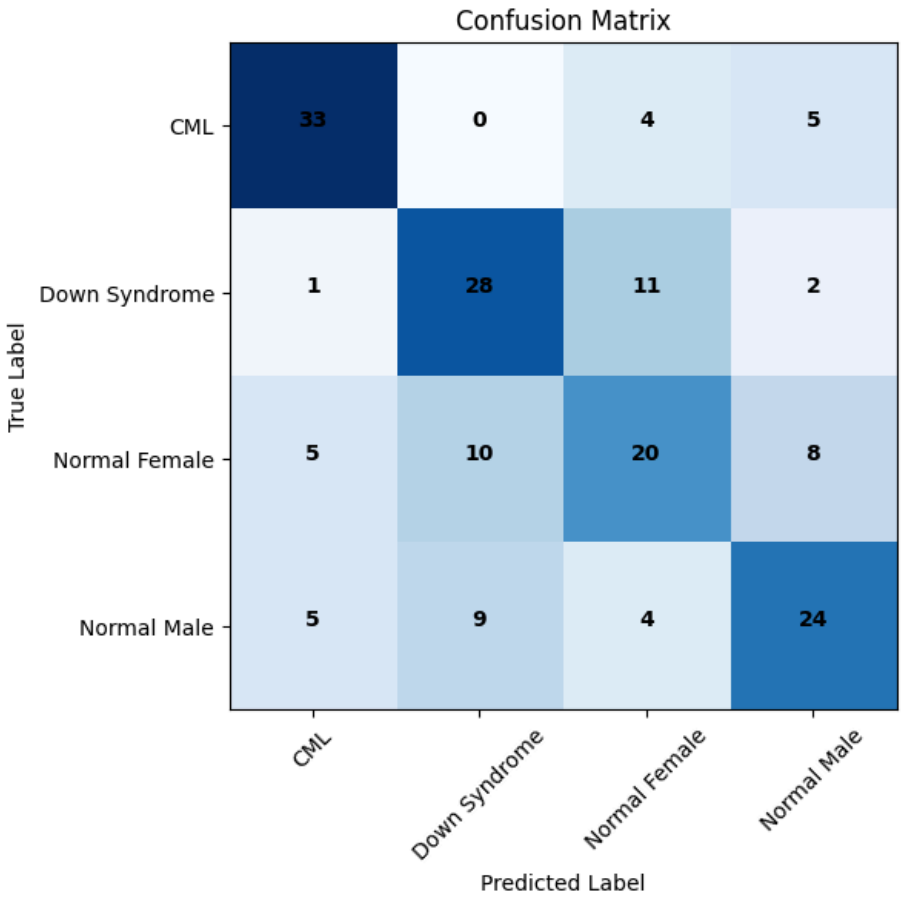


Confusion matrix for Naive Bayes

	precision	recall	f1-score	support
CML	0.50	0.68	0.58	38
Down Syndrome	0.60	0.68	0.64	38
Normal Female	0.78	0.64	0.70	44
Normal Male	0.69	0.53	0.60	47
accuracy			0.63	167
macro avg	0.64	0.63	0.63	167
weighted avg	0.65	0.63	0.63	167

Performance metrices for Naive Bayes

InceptionV3, a convolutional neural network model, achieved an accuracy of 62%. While it performed reasonably well, its accuracy was lower compared to 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.

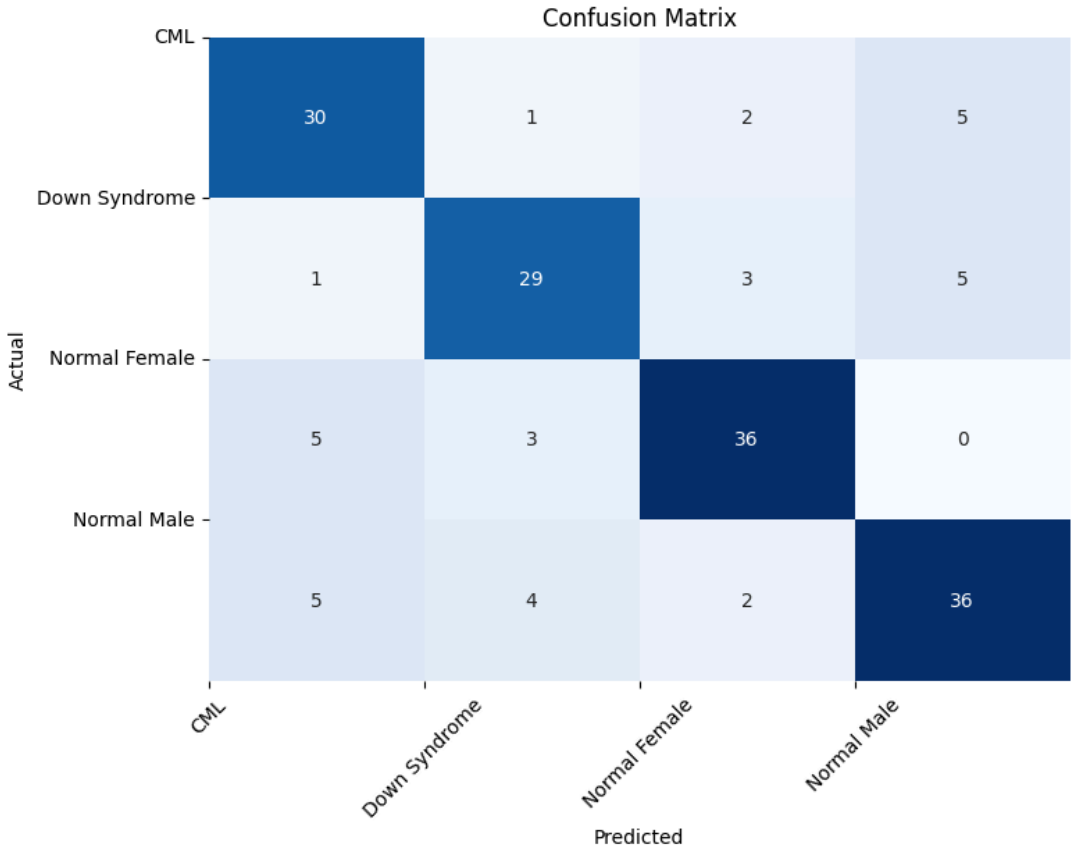


Confusion matrix for InceptionV3

	precision	recall	f1-score	support
CML	0.75	0.79	0.77	42
Down Syndrome	0.60	0.67	0.63	42
Normal Female	0.51	0.47	0.49	43
Normal Male	0.62	0.57	0.59	42
accuracy			0.62	169
macro avg	0.62	0.62	0.62	169
weighted avg	0.62	0.62	0.62	169

Performance metrices for InceptionV3

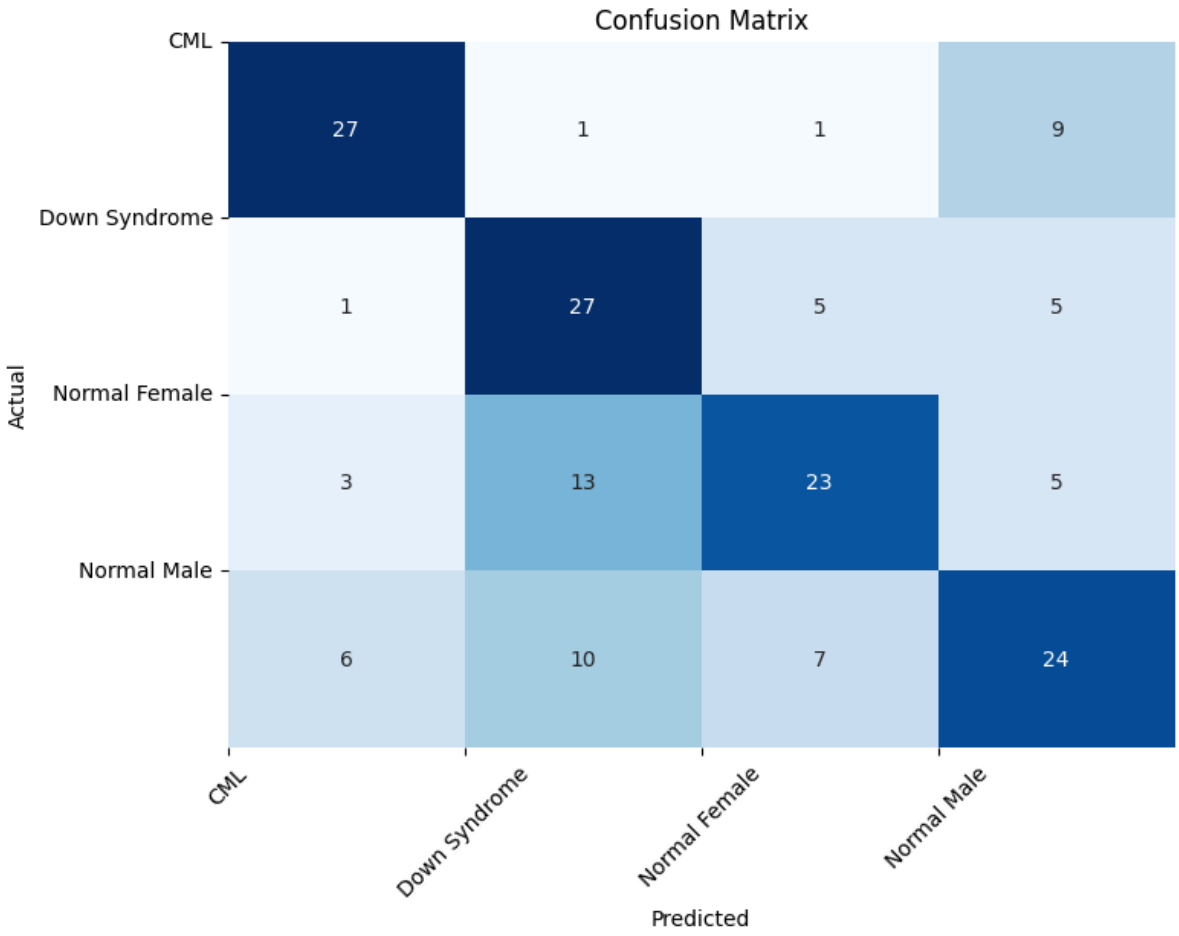
The hybrid models, XceptionBoost and IncepANN, achieved accuracies of 78% and 60% respectively. These models combine the feature extraction capabilities of InceptionV3 with other algorithms such as XGBoost and artificial neural networks. While they showed promising results, their accuracies were slightly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



Confusion matrix for XceptionBoost

	precision	recall	f1-score	support
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	0.77	47
accuracy			0.78	167
macro avg	0.78	0.78	0.78	167
weighted avg	0.79	0.78	0.78	167

Performance metrices for XceptionBoost



Confusion matrix for IncepANN

	precision	recall	f1-score	support
CML	0.73	0.71	0.72	38
Down Syndrome	0.53	0.71	0.61	38
Normal Female	0.64	0.52	0.57	44
Normal Male	0.56	0.51	0.53	47
accuracy			0.60	167
macro avg	0.61	0.61	0.61	167
weighted avg	0.61	0.60	0.60	167

Performance metrices for IncepANN

Conclusion

In conclusion, this project represents a significant step towards automated chromosomal disorder identification from karyotype images using machine learning models. The development of this system was driven by the lack of publicly available datasets and the time-consuming nature of traditional methods. By collaborating with the Kshema Centre for Genetic Services and collecting a comprehensive dataset, we were able to overcome these challenges and build a robust knowledge base from scratch.

By conducting a comparative analysis of different models, including XGBoost, SVM, Random Forest, Naive Bayes, and InceptionV3, we have shed light on the potential of these models in accurately classifying chromosomal disorders. The results revealed that XGBoost and Random Forest emerged as the top-performing models exhibiting superior performance in accurately classifying chromosomal disorders, surpassing traditional methods and even deep learning approaches., achieving impressive accuracies of 91% and 85% respectively. These findings demonstrate the effectiveness of ensemble-based approaches and decision tree algorithms in tackling the complexity of chromosomal disorder classification.

While the hybrid models, utilizing InceptionV3 for feature extraction and feeding into XGBoost and ANN, showed promising results, further improvements and fine-tuning may be required to unleash their full potential. Additionally, the availability of larger and more diverse datasets can contribute to enhancing the robustness and generalization capabilities of the models.

This project not only presents a powerful solution for chromosomal disorder classification but also sets the stage for future advancements in medical image analysis and diagnosis. The integration of such automated systems with clinical workflows holds great promise in improving patient outcomes, reducing human errors, and enhancing the overall efficiency of genetic disorder diagnosis.

Our findings highlight the potential of machine learning models in revolutionizing chromosomal disorder identification. The developed system serves as a stepping stone towards automated and accurate diagnoses, paving the way for early interventions and improved patient care. As further research and advancements continue, we anticipate even greater breakthroughs in the field of medical image analysis and genetic disorder detection.

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.