Capstone Project

DeepSkin: AI-Based Cancer Classification

Name: 1. Nandan Patel

2. Nirjari Bhatt

3. Krisha Mangukiya

4. Devansh Shrivastav

Contents

1. Project Statement

Skin cancer is one of the most prevalent types of cancer worldwide, affecting millions of people annually. Early detection plays a crucial role in effective treatment and improving survival rates. However, traditional diagnostic approaches primarily rely on clinical examination and biopsy, which can be time-consuming, expensive, and dependent on the experience of medical professionals. The manual classification of skin lesions often leads to varying accuracy levels, making early detection less reliable.

This project aims to address these challenges by developing an automated skin cancer classification system using deep learning. By leveraging Convolutional Neural Networks (CNNs), the system is capable of classifying different types of skin cancer with high accuracy. This AI-driven solution seeks to assist dermatologists and medical practitioners in providing timely and precise diagnoses, reducing the dependency on subjective evaluation, and making early intervention more accessible.

2. Project Overview

This project focuses on utilizing deep learning techniques to classify different types of skin cancer based on medical images. The dataset used for training and evaluation is stored in Google Drive and processed in Google Colab. The images are categorized into distinct skin cancer types, allowing the model to learn patterns and features for classification.

The workflow of the project consists of several critical stages:

1. **Data Handling**: Organizing, loading, and analyzing the dataset.
2. **Preprocessing**: Preparing images for training through augmentation and normalization.
3. **Model Training**: Developing and training a CNN model for classification.
4. **Model Evaluation**: Assessing the model's accuracy and performance using validation data.
5. **Model Deployment**: Saving the trained model for future use in real-world applications.

By following these structured phases, the project ensures a systematic approach to solving the problem of skin cancer classification using artificial intelligence.

3. Solution Offered

The proposed solution is an AI-driven skin cancer classification model trained on medical images. The deep learning pipeline includes the following components:

1. Data Handling

- The dataset, consisting of skin lesion images labeled by category, is stored in Google Drive.
- Google Colab is used to mount the drive and access the dataset.
- The number of images per class is analyzed to maintain a balanced dataset and avoid bias in the model.

Dataset Summary

Class Label	Class Name	Train	Validation	Test
6	seborrheic keratosis	1591	382	527
5	pigmented benign keratosis	1578	445	477
3	melanoma	1599	406	495
8	vascular lesion	1610	404	486
0	actinic keratosis	1596	405	499
1	basal cell carcinoma	1620	398	482
2	dermatofibroma	1632	369	499
4	nevus	1552	411	537
7	squamous cell carcinoma	1622	380	498
Total		14400	3600	4500

Figure 1 Data Summary

2. Preprocessing

- Images are read from the dataset directories.
- Labels are assigned based on folder names, ensuring consistency.
- Data is split into training and validation sets to optimize model learning.
- Image augmentation techniques (such as rotation, flipping, and scaling) are applied to improve model generalization.
- Normalization is performed to standardize pixel values and enhance computational efficiency.

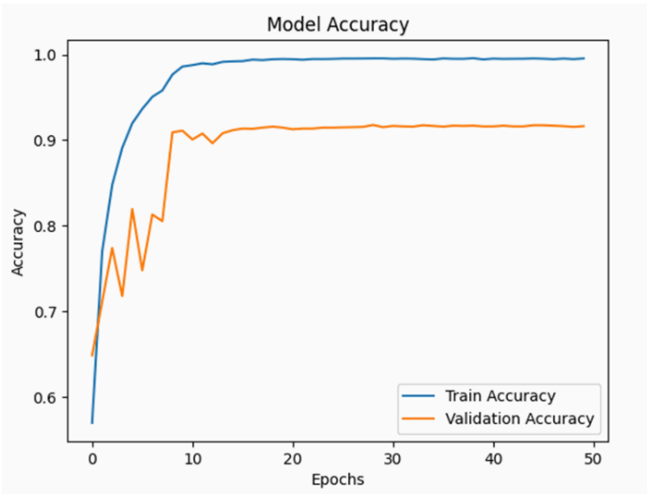
3. Model Training

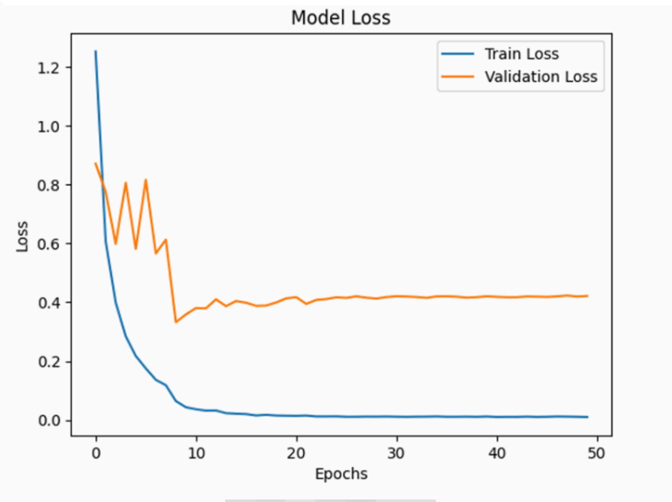
- A deep learning model based on Convolutional Neural Networks (CNNs) is built using TensorFlow and Keras.

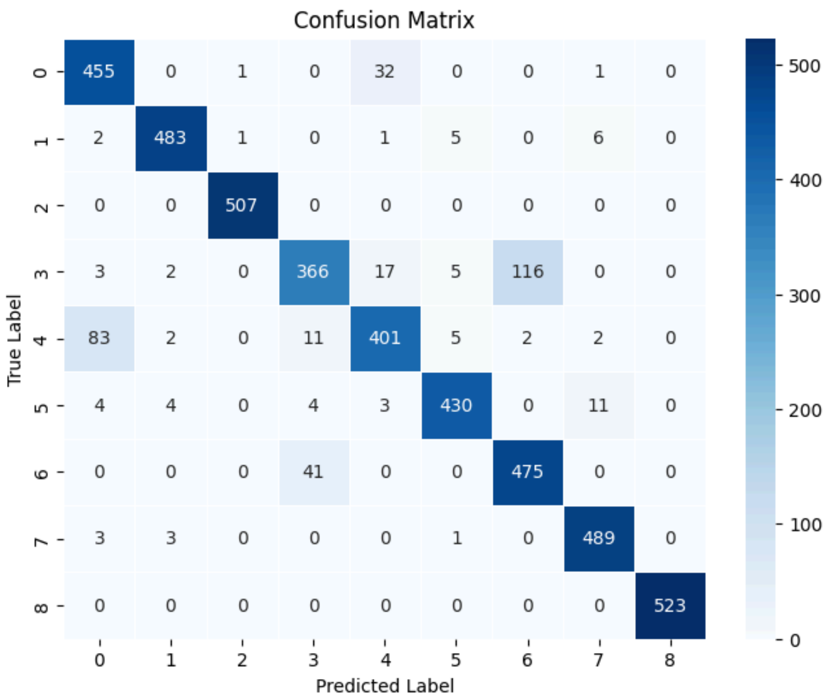
- The model architecture is designed with multiple convolutional layers, activation functions, pooling layers, and fully connected layers to extract and classify features efficiently.
- The model is compiled with a suitable optimizer (such as Adam) and a categorical cross-entropy loss function to improve classification accuracy.
- The training process is monitored using accuracy and loss curves.
- A checkpoint mechanism is implemented to save the best-performing model.

4. Model Evaluation

- The trained model is validated using unseen data.
- Performance metrics such as accuracy, precision, recall, F1-score, and confusion matrices are analyzed to assess effectiveness.
- Overfitting is mitigated by evaluating loss trends and using techniques such as dropout layers and regularization.







5. Model Deployment

- The trained model is saved in HDF5 format (`model.h5`) for real-world application.
- Training history, including loss and accuracy trends, is stored in a JSON file (`training_history.json`) for reference and further analysis.
- The model is ready for integration into clinical settings or telemedicine platforms to aid in dermatological diagnosis.

4. Who is the End Users?

The developed skin cancer classification system is designed for various end users, including:

- **Dermatologists and Medical Professionals**: To assist in diagnosing skin cancer quickly and accurately, serving as a secondary diagnostic tool.
- **Hospitals and Clinics**: To integrate AI-powered diagnostic tools into medical imaging workflows, enhancing efficiency and reducing misdiagnosis.
- **Medical Researchers**: To utilize the trained model for further research in dermatological AI and medical image processing.
- **Telemedicine Platforms**: To offer remote skin cancer detection services, enabling accessibility for patients in underserved areas.
- **Patients and General Users**: To enable early self-assessment before seeking professional consultation, promoting proactive healthcare awareness.

5. Technologies Used to Solve this Problem

The successful implementation of this project relies on a combination of technologies and tools, including:

- **Google Colab**: A cloud-based platform for training and processing deep learning models efficiently.
- **Python**: The primary programming language used for data handling, model training, and evaluation.
- **TensorFlow/Keras**: Deep learning frameworks used to build and train the CNN model.
- **OpenCV**: For image processing, enhancement, and augmentation.
- **OS & Pandas**: For file handling and efficient data manipulation.

- **Matplotlib & Seaborn**: For visualizing training progress, loss curves, and evaluation metrics.
- **NumPy & Scikit-Learn**: For numerical operations and statistical analysis of model performance.

6. Conclusion

This project successfully develops a deep learning-based skin cancer classification system using CNNs. By leveraging AI techniques, the model enhances the efficiency and accuracy of skin cancer diagnosis, providing a robust tool for medical professionals and researchers. The integration of deep learning in dermatology has the potential to revolutionize early detection methods, ensuring timely intervention and improved patient outcomes.

The trained model can be further improved with larger datasets, additional preprocessing techniques, and more advanced architectures. Future work may involve real-time implementation in hospitals, developing mobile applications for instant classification, and enhancing model robustness with additional medical imaging datasets.

Through this project, we demonstrate how artificial intelligence can contribute to the healthcare industry, bridging the gap between medical expertise and technology-driven solutions for better diagnostic capabilities.