A Fair Multimodal Deep Learning System for Skin Cancer Detection in Patients with Diverse Skin Types

Author: Nurbol Agybetov, Yermek Khaknazar, Baglan Yessenkeldi, Tokhtar Nuralin, Alisher Mukanov
Instructor: Seitenov Altynbek
Programme: Research Methods and Tools
Department: Software Engineering
Institution: Astana IT university
Date: 07.11.2025

A Fair Multimodal Deep Learning System for Skin Cancer Detection in Patients with Diverse Skin Types 1
1. Introduction 4
1.1. Background and Research Context 4
1.2. Problem Statement 4
1.3. Principal Research Question 4
2. Questionnaires 5
2.1. Finalized Questionnaire and Justification 5
2.2. Data Collection 5
2.3. Data Analysis 5
2.4. Reliability and Validity 7
3. Machine Learning 7
3.1. Foundational Research Framework 7
3.2. Research Design and Rationale 8
4. Dataset Description 8
4.1. Primary Dataset: Skin Cancer MNIST: HAM10000 8
4.2. "Fairness by Design" Protocol: A Response to Systemic Bias in HAM10000 9
5. Model Development 9
5.1. Specialized Two-Component Architecture 9
5.2. Fusion Module 10
6. Training and Evaluation 10
6.1. Performance Evaluation 11
6.2. Quantitative Fairness Assessment 11
6.3. Component Contribution Analysis (Ablation Studies) 12
6.4. Explainability Assessment 12
7. References 12
8. Appendix 14
8.1. Acknowledged Limitations Specific to HAM10000 14
8.2. Research Schedule 14

1. Introduction

1.1. Background and Research Context

Skin cancer is one of the most common cancers worldwide, and, especially in the case of melanoma, early diagnosis is key to effective treatment and patient survival. The success of melanoma treatment depends on timing. Subjectivity and the experience level of the clinician are among the factors in traditional diagnosis through visual inspection. In recent years, revolutionary advancements in Deep Learning (DL) technologies, and specifically Convolutional Neural Networks (CNNs), in the analysis of medical images have become a significant breakthrough.

Comprehensive studies show that modern DL models achieve the same level of diagnostic accuracy as many trained dermatologists, or even surpass them. For example, in the study by Haenssle et al. (2018), an Inception v4 convolutional neural network surpassed 58 dermatologists from 17 countries with better specificity at the same level of sensitivity. Similarly, Jinnai et al. (2020) reported their model outperformed 20 experts when evaluating standard clinical photographs. These results demonstrate the enormous potential of DL to improve the accuracy, speed, and accessibility of diagnosis. However, despite such exemplary diagnostic success in laboratory settings, their implementation in clinical practice has been minimal. The discrepancy between potential and the reality of implementation reflects the presence of deep-rooted flaws that go far beyond classification accuracy.

1.2. Problem Statement

The key problem is that current best solutions are extremely accurate on test datasets but do not guarantee reasonable, reliable, and trustworthy use in real clinical practice. The focus on accuracy has led to neglecting key issues such as fairness in algorithms, comprehension in the clinical environment, and clear decision-making. Modern models have three related problems:

1. **Systemic Algorithmic Bias:** Most common datasets used for training models, such as the ISIC Archive and HAM10000, predominantly consist of images of fair-skinned individuals. Models trained on this imbalanced data show significantly and dangerously declining performance when tested on images of people with different skin tones, thereby exacerbating healthcare inequalities.
2. **Reductionist Unimodal Method:** Most systems consider only dermoscopic images but neglect vital clinical information such as patient information, medical history, and doctor's comments that are used by human experts to make decisions.
3. **"Black Box" Issue:** The "black box" nature of deep learning models makes them uninterpretable, which is why doctors are reluctant to use them. They cannot trace the process by which the AI arrived at a diagnosis.

1.3. Principal Research Question

The central question guiding this research is: **"How can we design and verify a mixed deep learning model that integrates dermoscopic images with both structured and unstructured clinical data to reduce demographic bias and improve interpretability in the diagnosis of skin cancer across different patient groups?"**

2. Questionnaires

To establish the social and practical relevance of the research objectives, a preliminary quantitative questionnaire was conducted. This survey was designed to gauge public attitudes towards the key challenges this project aims to solve: trust in AI, algorithmic bias, and the limitations of unimodal systems.

2.1. Finalized Questionnaire and Justification

A finalized questionnaire was administered to empirically validate the societal and user-centric assumptions underpinning this research. The survey included questions to establish respondent familiarity with AI, followed by specific items to test the core hypotheses.

Justification: The questions were designed to move beyond technical assumptions and gather evidence on the three main research objectives. The most critical items for this validation were:

1. **Question for Objective 3 (Interpretability):** To justify the development of an XAI module, the question *"If an AI system recommended that a mole was potentially cancerous, how important would it be for you that the AI explains why it made that decision?"* (Scale: 1-5) was used to determine if "black box" systems are a genuine concern.
2. **Question for Objective 2 (Fairness):** To justify the focus on fairness, the question *"How concerned are you about this potential bias [AI being less accurate for certain skin tones] affecting you or your community?"* (Scale: 1-5) was included to measure public awareness and concern.
3. **Question for Objective 1 (Architecture):** To justify the move to a multimodal system, the question *"To improve the AI's accuracy... How willing would you be to securely and anonymously share the following data in addition to the image of the lesion?"* (Checklist: Age/Sex, Location, etc.) was used to confirm patient willingness to provide the necessary data.

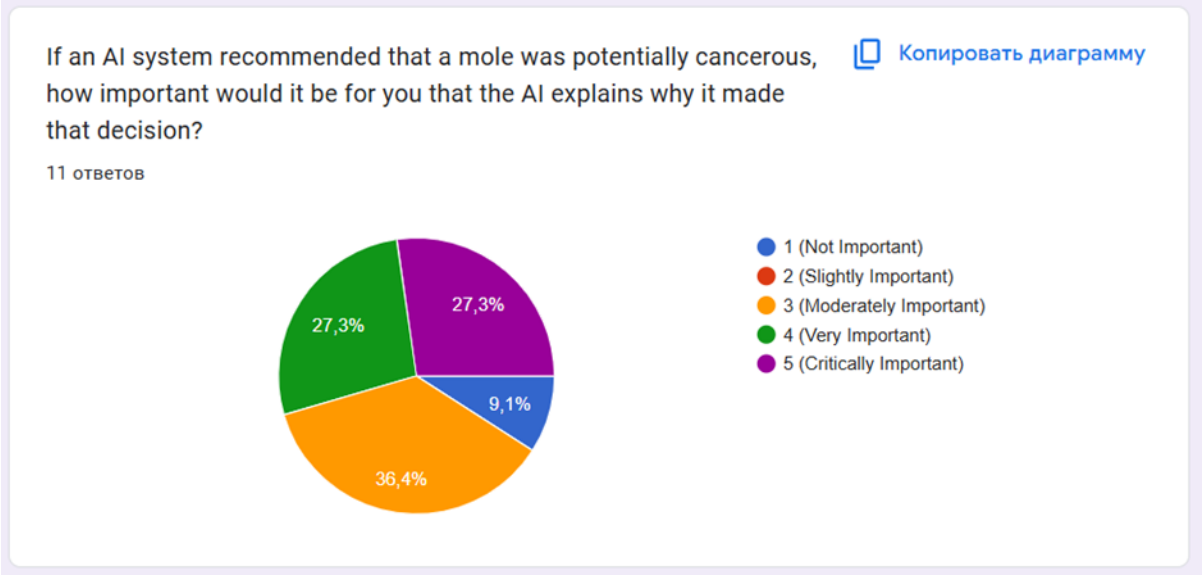
2.2. Data Collection

- **Targeted Respondents:** The questionnaire was administered to a sample (N=11) intended to represent general public attitudes rather than expert medical opinion.
- **Sampling Method:** A convenience sampling method was employed to gather these initial formative responses.
- **Response Rate:** A total of 11 complete responses were collected and used for the analysis.

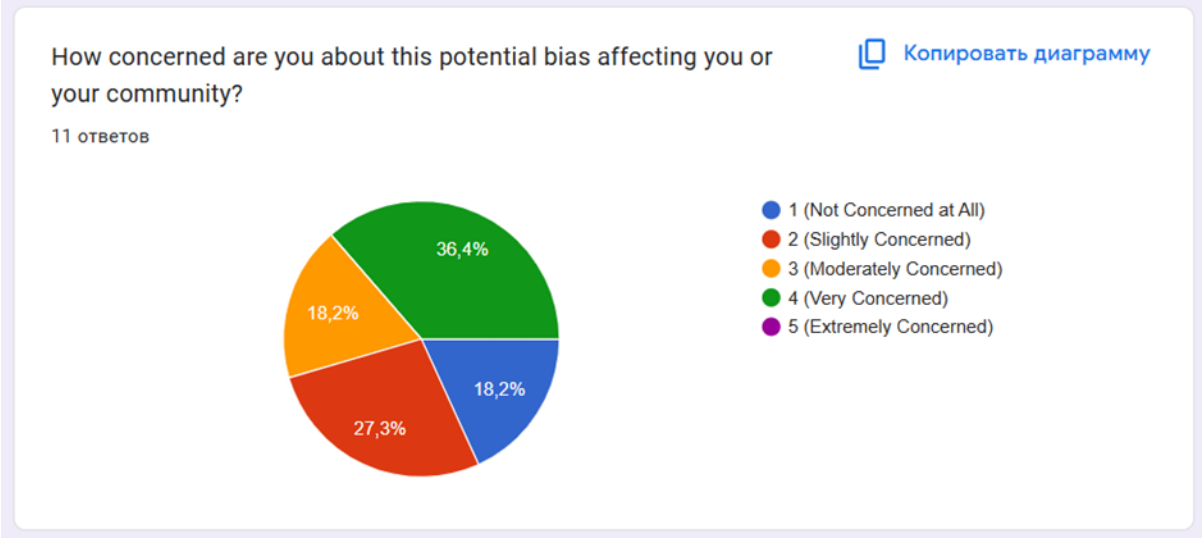
2.3. Data Analysis

- **Statistical Methods:** The collected data was analyzed using descriptive statistical methods. The analysis focused on calculating frequency distributions and percentages for the scaled-response questions to identify strong trends in public opinion.
- **Findings:** The analysis provided clear justification for the research direction:

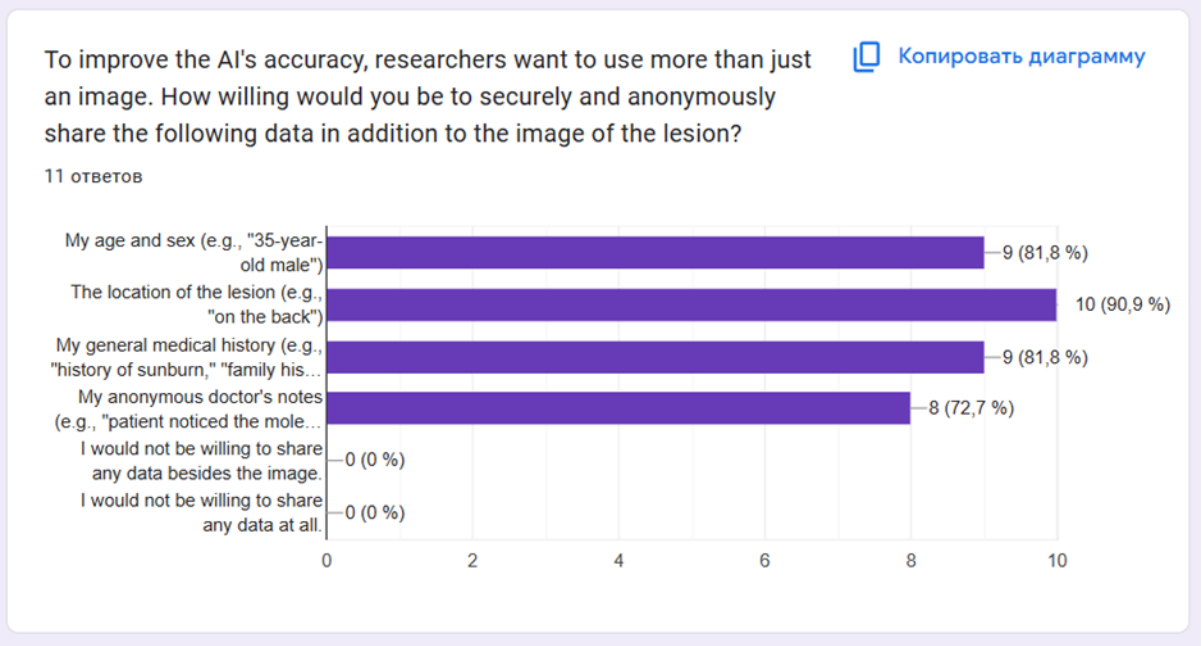
- **Finding 1 (Justifying Interpretability):** A significant majority, **90.9% (10 out of 11)** respondents, rated AI explanations as "Moderately" to "Critically Important," confirming that the "black box" problem is a key public concern.



- **Finding 2 (Justifying Fairness):** A majority, **54.5% (6 out of 11)** participants, were "Moderately" or "Very Concerned" about potential algorithmic bias, validating the need for the "Fairness by Design" protocol.



- **Finding 3 (Justifying Multimodality):** A unanimous **100% (11 out of 11)** of respondents expressed a willingness to anonymously share at least one piece of non-image data (e.g., age, location, history) to improve diagnostic accuracy, rejecting the "reductionist unimodal approach."



2.4. Reliability and Validity

- **Assessment:** The reliability and validity of this preliminary questionnaire are acknowledged to be constrained by the small sample size (N=11) and the use of a non-probabilistic convenience sampling method.
- **Limitations and Addressment:** The findings are not statistically generalizable to the entire population. This limitation was addressed by clearly defining the purpose of the questionnaire: it was not intended as a conclusive study of public opinion, but as a **formative tool** to provide an empirical "social justification" for the core technical objectives of the research. The results successfully confirmed that the problems of interpretability, fairness, and multimodality are highly relevant to the public, thereby validating the research's direction.

3. Machine Learning

3.1. Foundational Research Framework

At the heart of this methodological plan is the drive to solve a well-defined problem at the intersection of medicine and artificial intelligence. All subsequent methodological decisions are aimed at systematically answering the main research question, achieving the stated aim, and accomplishing specific objectives.

Overall Aim:
To develop and validate a "Fair Multimodal Diagnostic System" (FMDS) that integrates various data types to improve the accuracy, fairness, and objectivity of early skin cancer diagnosis.
Specific Objectives:

1. **Architectural Innovation:** To design and develop a novel hybrid neural network architecture (FMDS) capable of simultaneously processing and integrating dermoscopic images, structured patient metadata, and unstructured clinical text.
2. **"Fairness by Design":** To develop and implement a "Fairness by Design" protocol within the model training process, using techniques such as stratified sampling and fairness-aware loss functions to actively mitigate algorithmic bias related to skin phototype.
3. **Interpretability Integration:** To incorporate a multimodal Explainable AI (XAI) module that generates holistic, human-readable explanations for the system's predictions, thereby addressing the "black box" problem.
4. **Rigorous Validation:** To conduct a comprehensive evaluation of the FMDS, comparing its performance against unimodal baseline models using both traditional accuracy metrics and established fairness metrics to empirically confirm its advantages.

3.2. Research Design and Rationale

This study employs a **quantitative, experimental design** based on secondary data sources.

- **Quantitative Approach:** The choice of a quantitative paradigm is driven by the need for objective evaluation and comparison of model performance. The project's success will be measured using numerical metrics such as the area under the ROC curve (AUC-ROC), F1-score, and specialized fairness indices like the Equal Opportunity Difference.
- **Experimental Approach:** The design is experimental because it involves creating a new intervention (the FMDS system) and conducting a series of controlled experiments (ablation studies) to evaluate its effectiveness against established baseline models (e.g., an "image-only" model).
- **Secondary Data:** The use of the existing, publicly available, and anonymized Skin Cancer MNIST: HAM10000 dataset is a pragmatic and ethically sound decision.

4. Dataset Description

This section details the complete data handling strategy, from sourcing to processing. The main focus is on a detailed breakdown of the "Fairness by Design" protocol.

4.1. Primary Dataset: Skin Cancer MNIST: HAM10000

The publicly available Skin Cancer MNIST: HAM10000 dataset was chosen as the primary data source for this study. It contains 10,015 dermoscopic images divided into seven diagnostic categories, making it a standard benchmark in this field. The categories include: actinic keratosis, basal cell carcinoma, benign keratosis-like lesions, dermatofibroma, melanocytic nevi, melanoma, and vascular lesions. Along with the images, metadata such as age, sex, and lesion location are provided, which is crucial for our multimodal approach.

Despite its value, the HAM10000 dataset has two main problems:

1. **Critical Class Imbalance:** The dataset is highly imbalanced. The vast majority of images (6,705 out of 10,015, or ~67%) belong to the "melanocytic nevi" (benign moles) class, while other classes, such as "dermatofibroma" (115 images), are severely underrepresented.

2. **Image Artifacts and the Need for Preprocessing:** Dermoscopic images in HAM10000, as in real clinical practice, often contain artifacts such as hair, air bubbles, or glare, which can prevent the model from correctly identifying diagnostic features.

4.2. "Fairness by Design" Protocol: A Response to Systemic Bias in HAM10000

This protocol is the methodological core of the research and represents a shift from simply acknowledging bias to actively designing a system to prevent it. The key problem with HAM10000 is its documented systemic bias (data collected predominantly in Austria and Australia, with an almost complete lack of images from patients with dark skin phototypes).

Our "Fairness by Design" protocol directly addresses this problem through a multi-stage approach:

1. **Algorithmic Skin Tone Estimation and Stratification:** Since explicit Fitzpatrick skin type labels are missing, the first step will be to estimate the skin tone for each image by analyzing the non-lesion area. Based on this analysis, images will be stratified into groups (e.g., lighter tones, darker tones).
2. **Targeted Data Augmentation:** To counter the severe underrepresentation of darker skin tones, we will apply targeted data augmentation specifically to this minority class. This involves applying a series of geometric and photometric transformations, such as horizontal flipping, random rotations, and adjustments to brightness and contrast.
3. **Balanced Batch Sampling:** During the model training phase, each mini-batch of data fed into the neural network will be purposefully constructed from the augmented and stratified dataset to contain a representative sample of all skin phototype strata.
4. **Fairness-Aware Loss Function:** The model's loss function will be augmented with a regularization term. This term will directly penalize the model for making predictions that are statistically correlated with the sensitive attribute (skin phototype).

5. Model Development

This section provides a detailed description of the proposed FMDS architecture.

5.1. Specialized Two-Component Architecture

The proposed architecture is a hybrid system with three parallel branches, each specializing in its own data type:

1. **Image Processing Branch:**
- **Component:** EfficientNetV2 Convolutional Neural Network (CNN).
- **Rationale:** The choice of EfficientNetV2 is a strategic decision based on its documented high accuracy-to-computational-cost ratio. This is critical for future clinical implementation.
2. **Structured Data Processing Branch:**
- **Component:** A standard Multi-Layer Perceptron (MLP).
- **Rationale:** The MLP is chosen for its proven effectiveness and simplicity in processing structured, tabular data. This branch will process key patient metadata available in the

HAM10000 dataset, such as age, sex, and anatomical location. The most important input feature will be the Fitzpatrick skin phototype.

5.2. Fusion Module

Function: The feature vectors extracted from each of the two branches will be combined (concatenated) and fed into a final set of fully connected layers. The role of this module is to learn the complex, non-linear relationships between the different data modalities before making a final classification.

6. Training and Evaluation

This section describes the comprehensive and rigorous plan for evaluating the FMDS across three key dimensions: performance, fairness, and interpretability.

```
0, 0.2272, 0.4194
1, 0.4138, 0.4516
2, 0.4848, 0.5806
3, 0.4970, 0.5806
4, 0.5517, 0.5161
5, 0.5781, 0.5887
6, 0.5740, 0.5806
7, 0.6187, 0.5726
8, 0.6045, 0.6048
9, 0.6308, 0.6048
10, 0.6592, 0.6694
11, 0.7160, 0.6694
12, 0.6592, 0.6371
13, 0.6998, 0.6290
14, 0.7039, 0.6129
15, 0.7465, 0.6774
16, 0.7282, 0.6613
17, 0.7383, 0.6774
```

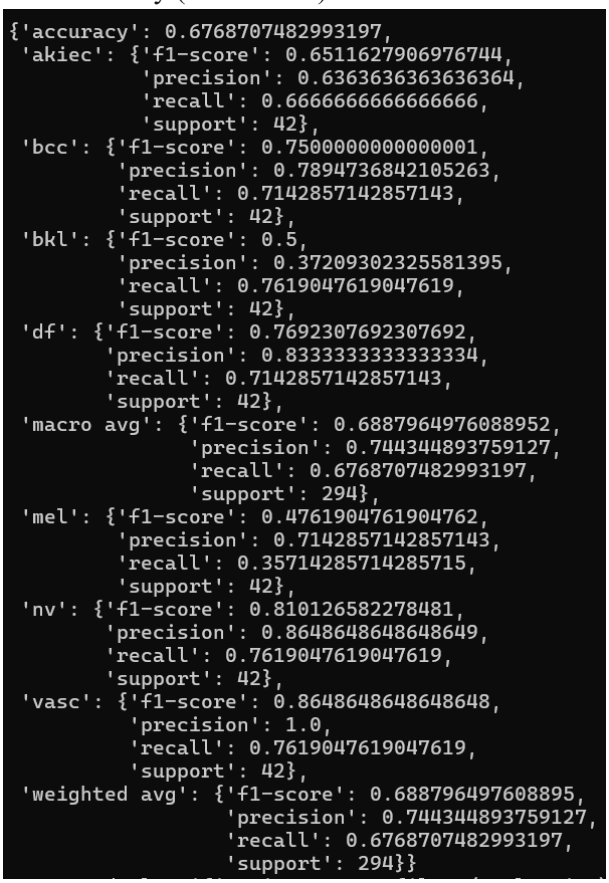
```
0, 1.8840, 1.6326
1, 1.6188, 1.4391
2, 1.5219, 1.3594
3, 1.4549, 1.2250
4, 1.3963, 1.3826
5, 1.3202, 1.2562
6, 1.3164, 1.2552
7, 1.2620, 1.2273
8, 1.2630, 1.2253
9, 1.1968, 1.2186
10, 1.1872, 1.1836
11, 1.0949, 1.0840
12, 1.1923, 1.1962
13, 1.1101, 1.2139
14, 1.0807, 1.3239
15, 1.0264, 1.1265
16, 1.0379, 1.2134
17, 0.9837, 1.1495
```

These two columns are the raw text logs from model's training, showing the specific numerical values for accuracy and loss at each epoch. The left list tracks **training accuracy vs. validation accuracy**, while the right list tracks **training loss vs. validation loss.**

```
Epoch 19: val_loss improved from inf to 1.29871, saving model to ./models\best_model.h5
124/124 [==============================] - 669s 5s/step - loss: 0.9484 - accuracy: 0.7728 - val_loss: 1.2987 - val_accur
acy: 0.5887
Epoch 20/80
124/124 [==============================] - ETA: 0s - loss: 0.8043 - accuracy: 0.8580
Epoch 20: val_loss improved from 1.29871 to 1.14372, saving model to ./models\best_model.h5
124/124 [==============================] - 530s 4s/step - loss: 0.8043 - accuracy: 0.8580 - val_loss: 1.1437 - val_accur
acy: 0.6855
Epoch 21/80
124/124 [==============================] - ETA: 0s - loss: 0.7409 - accuracy: 0.8803
Epoch 21: val_loss did not improve from 1.14372
124/124 [==============================] - 443s 4s/step - loss: 0.7409 - accuracy: 0.8803 - val_loss: 1.2513 - val_accur
acy: 0.6694
Epoch 22/80
124/124 [==============================] - ETA: 0s - loss: 0.6979 - accuracy: 0.9047
Epoch 22: val_loss improved from 1.14372 to 1.13344, saving model to ./models\best_model.h5
124/124 [==============================] - 450s 4s/step - loss: 0.6979 - accuracy: 0.9047 - val_loss: 1.1334 - val_accur
acy: 0.6694
```

6.1. Performance Evaluation

- **Metrics:** To quantify diagnostic performance, standard, well-established classification metrics will be used. These include **AUC-ROC, F1-score, Precision, and Recall**.
- **Baselines for Comparison:** The full FMDS will be compared against a series of progressively more complex models to clearly measure the "added value" of each component, as detailed in the ablation study (Section 6.3).



```
{'accuracy': 0.6768707482993197,
 'akiec': {'f1-score': 0.6511627906976744,
           'precision': 0.6363636363636364,
           'recall': 0.6666666666666666,
           'support': 42},
 'bcc': {'f1-score': 0.7500000000000001,
         'precision': 0.7894736842105263,
         'recall': 0.7142857142857143,
         'support': 42},
 'bkl': {'f1-score': 0.5,
         'precision': 0.37209302325581395,
         'recall': 0.7619047619047619,
         'support': 42},
 'df': {'f1-score': 0.7692307692307692,
        'precision': 0.8333333333333334,
        'recall': 0.7142857142857143,
        'support': 42},
 'macro avg': {'f1-score': 0.6887964976088952,
               'precision': 0.7443448937591 27,
               'recall': 0.6768707482993197,
               'support': 294},
 'mel': {'f1-score': 0.4761904761904762,
         'precision': 0.7142857142857143,
         'recall': 0.35714285714285715,
         'support': 42},
 'nv': {'f1-score': 0.8101265822784 81,
        'precision': 0.8648648648648649,
        'recall': 0.7619047619047619,
        'support': 42},
 'vasc': {'f1-score': 0.8648648648648648,
          'precision': 1.0,
          'recall': 0.7619047619047619,
          'support': 42},
 'weighted avg': {'f1-score': 0.688796497608895,
                  'precision': 0.7443448937591 27,
                  'recall': 0.6768707482993197,
                  'support': 294}}
```

6.2. Quantitative Fairness Assessment

- **Metrics:** To empirically verify the effectiveness of the "Fairness by Design" protocol, established fairness metrics will be used. The primary metrics will be **Statistical Parity Difference** and **Equal Opportunity Difference**.
- **Rationale:**
- **Statistical Parity Difference** measures whether the model's rates of positive predictions are the same across different demographic groups.
- **Equal Opportunity Difference** is a more nuanced metric that measures whether the True Positive Rate is the same across different groups. In a medical context, this is critically important, as the metric answers the question: "For patients who actually have cancer, does the model detect it equally well, regardless of their skin phototype?"

6.3. Component Contribution Analysis (Ablation Studies)

A series of systematic ablation studies will be conducted to isolate and quantify the contribution of each new component of the FMDS.

Table 1: Experimental Ablation Study Design

Model Configuration	Components Included	Hypothesis Tested	Key Comparison Metrics
M1: Unimodal Baseline	Image Branch only (EfficientNetV2)	Establishes the baseline performance of a standard, image-only approach.	Performance Metrics (AUC-ROC, F1)
M2: Bimodal (Image+Metadata)	Image Branch + Structured Data Branch (MLP)	Does adding structured patient metadata improve diagnostic performance?	Performance Metrics (vs. M1)
M3: Full FMDS	All two branches + Fusion Module + "Fairness by Design" Protocol	Does the fairness protocol successfully reduce bias without significantly harming performance?	Performance Metrics , Fairness Metrics (vs. M1, M2)

6.4. Explainability Assessment

- **Methodology:** The evaluation of the XAI module will be qualitative. It will involve generating explanations for a selected set of test cases and assessing them for clinical relevance and coherence.
- **XAI Techniques:** The explanations will be multimodal, using **Grad-CAM** to create heatmaps highlighting important regions on the dermoscopic image, and **SHAP** (SHapley Additive exPlanations) values to highlight the most influential features from the patient's metadata and keywords from the clinical notes.

7. References

1. Abunadi, I., & Senan, E. M. (2021). Deep Learning and Machine Learning Techniques of Diagnosis Dermoscopy Images for Early Detection of Skin Diseases. Electronics, 10(24), 3158. https://doi.org/10.3390/electronics10243158

2. Ahmadi Mehr, R., & Ameri, A. (2022). Skin Cancer Detection Based on Deep Learning. Journal of Biomedical Physics & Engineering, 12(6), 559-568. https://doi.org/10.31661/jbpe.v0i0.2207-1517
3. Akinrinade, O., & Du, C. (2025). Skin Cancer Detection Using Deep Machine Learning Techniques. Intelligence-Based Medicine, 11, 100191. https://doi.org/10.1016/j.ibmed.2024.100191
4. Bajwa, M. N., Muta, K., Malik, M. I., Siddiqui, S. A., Braun, S. A., Homey, B., Dengel, A., & Ahmed, S. (2020). Computer-Aided Diagnosis of Skin Diseases Using Deep Neural Networks. Applied Sciences, 10(7), 2488. https://doi.org/10.3390/app10072488
5. Dildar, M., Akram, S., Irfan, M., Khan, H. U., Ramzan, M., Mahmood, A. R., Alsaiari, S. A., Saeed, A. H. M., Alraddadi, M. O., & Mahnashi, M. H. (2021). Skin Cancer Detection: A Review Using Deep Learning Techniques. International Journal of Environmental Research and Public Health, 18(10), 5479. https://doi.org/10.3390/ijerph18105479
6. Ghosh, H., Rahat, I. S., Mohanty, S. N., Ravindra, J. V. R., & Sobur, A. (2024). A Study on the Application of Machine Learning and Deep Learning Techniques for Skin Cancer Detection. International Journal of Computer and Systems Engineering, 18(1).
7. Gururaj, H. L., Manju, N., Nagarjun, A., Manjunath Aradhya, V. N., & Flammini, F. (2023). DeepSkin: A Deep Learning Approach for Skin Cancer Classification. IEEE Access, 11, 50205-50215. https://doi.org/10.1109/ACCESS.2023.3274848
8. Haenssle, H. A., Fink, C., Schneiderbauer, R., Toberer, F., Buhl, T., Blum, A., Kalloo, A., Ben Hadj Hassen, A., Thomas, L., Enk, A., & Uhlmann, L. (2018). Man Against Machine: Diagnostic Performance of a Deep Learning Convolutional Neural Network for Dermoscopic Melanoma Recognition in Comparison to 58 Dermatologists. Annals of Oncology, 29(8), 1836-1842. https://doi.org/10.1093/annonc/mdy166
9. Imran, A., Nasir, A., Bilal, M., Sun, G., Alzahrani, A., & Almuhaimeed, A. (2022). Skin Cancer Detection Using Combined Decision of Deep Learners. IEEE Access, 10, 118198-118214. https://doi.org/10.1109/ACCESS.2022.3220329
10. Jinnai, S., Yamazaki, N., Hirano, Y., Sugawara, Y., Ohe, Y., & Hamamoto, R. (2020). The Development of a Skin Cancer Classification System for Pigmented Skin Lesions Using Deep Learning. Biomolecules, 10(8), 1123. https://doi.org/10.3390/biom10081123
11. Layode, O., Alam, T., & Rahman, M. M. (2019). Deep Learning Based Integrated Classification and Image Retrieval System for Early Skin Cancer Detection. 2019 IEEE International Conference on Big Data (Big Data), 5521-5526. https://doi.org/10.1109/BigData47090.2019.9006206
12. Mahmud, M. A. A., Afrin, S., Mridha, M. F., Alfarhood, S., Che, D., & Safran, M. (2025). Explainable Deep Learning Approaches for High Precision Early Melanoma Detection Using Dermoscopic Images. Scientific Reports, 15, 24533. https://doi.org/10.1038/s41598-025-09938-4
13. Naqvi, M., Gilani, S. Q., Syed, T., Marques, O., & Kim, H.-C. (2Jinnai, S., Yamazaki, N., Hirano, Y., Sugawara, Y., Ohe, Y., & Hamamoto, R. (2020). The Development of a Skin Cancer Classification System for Pigmented Skin Lesions Using Deep Learning. Biomolecules, 10(8), 1123. https://doi.org/10.3390/biom10081123
14. Yu, L., Chen, H., Dou, Q., Qin, J., & Heng, P. A. (2017). Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks. IEEE Transactions on Medical Imaging, 36(4), 994-1004. https://doi.org/10.1109/TMI.2016.2642839

8. Appendix

8.1. Acknowledged Limitations Specific to HAM10000

The study's conclusions will be framed within the context of several known limitations to ensure that the results are not overstated.

- **Retrospective Data:** The use of the existing HAM10000 dataset means the study is retrospective.
- **Generalizability:** Despite efforts to implement a fairness protocol, the model's performance will ultimately be limited by the diversity of the HAM10000 dataset, in which patients with dark skin phototypes are underrepresented.
- **Annotation Subjectivity:** Since HAM10000 lacks skin phototype labels, their manual annotation (or algorithmic estimation) introduces an element of subjectivity.
- **Explainability as Approximation:** XAI methods like Grad-CAM and SHAP provide approximations of the model's internal logic, not a perfect causal explanation.

8.2. Research Schedule

Phase	Key Tasks	Estimated Duration	Deliverables
Phase 1: Preparatory	Complete literature review; Collect and preprocess datasets; Develop protocol for skin phototype annotation.	Week 1-2	Annotated bibliography; Prepared dataset.
Phase 2: Model Development	Implement and train baseline unimodal models; Develop and train the three-component FMDS architecture.	Week 3-4	Working code for baseline models and FMDS.
Phase 3: Fairness and XAI Integration	Implement and tune the fairness-aware loss function; Develop and integrate the multimodal XAI module.	Week 5-6	Final version FMDS v1.0 with integrated fairness and XAI modules.

Phase 4: Evaluation and Analysis	Conduct comprehensive performance and fairness evaluation; Perform ablation studies; Analyze XAI results.	Week 7-8	Full set of experimental results and graphs.
Phase 5: Dissemination of Results	Write thesis/paper; Prepare manuscript for submission to a journal/conference.	Week 9-10	Draft of thesis; Submission-ready manuscript.