Tele-Dermatology Platform in the Netherlands

Presented by:

- Anna-Maria Pervan: 6190676
- Florent Didascalou: 6337071
- Sijia Wang : 6349144
- Singharat Rattanaphan (Billy): 6311014
- Sonja Tang : 6345924

Maastricht University | Department of Advanced Computing Sciences

Section 1

Outcome Economy

Tele-dermatology service model (Business idea)

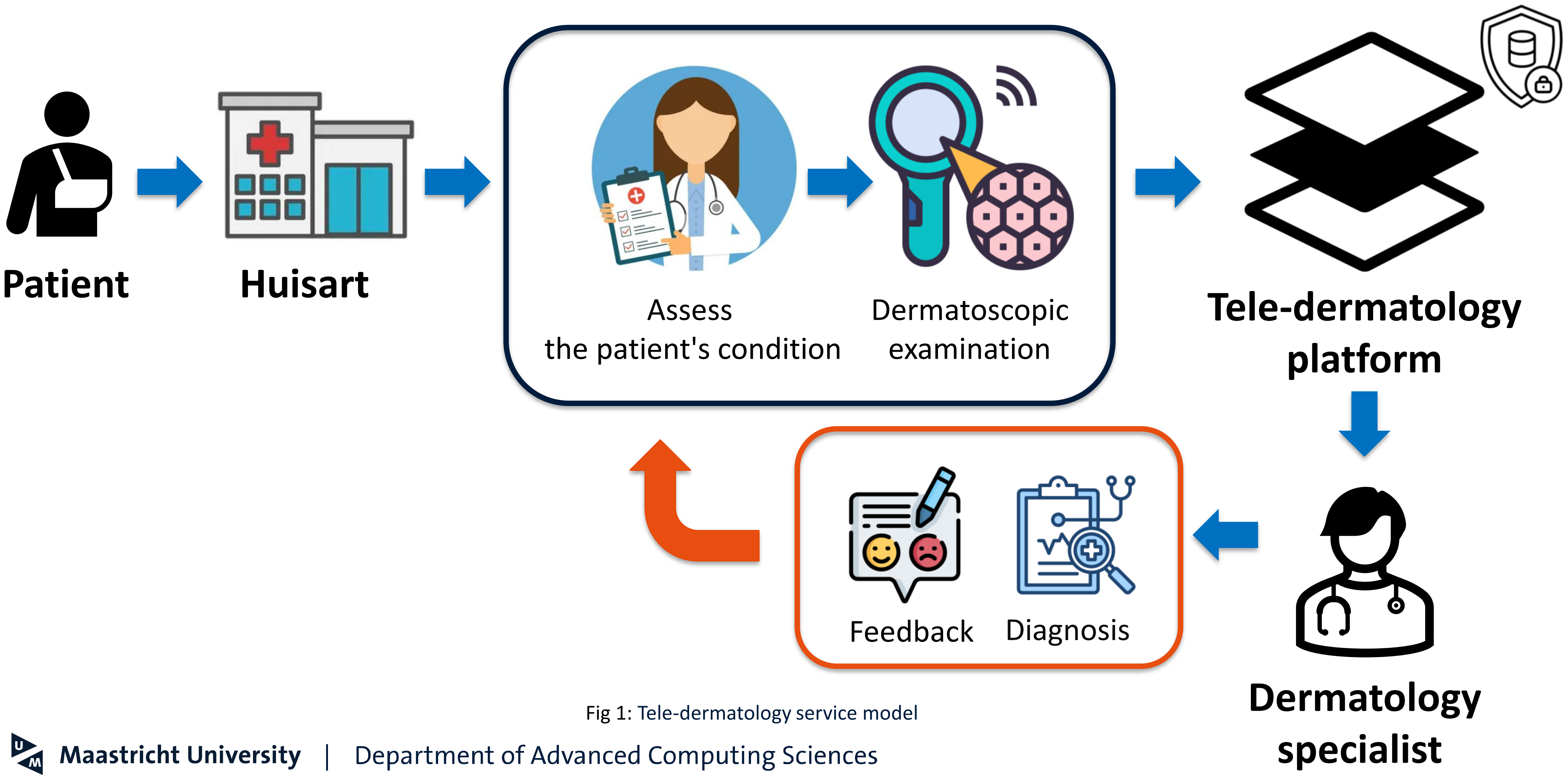
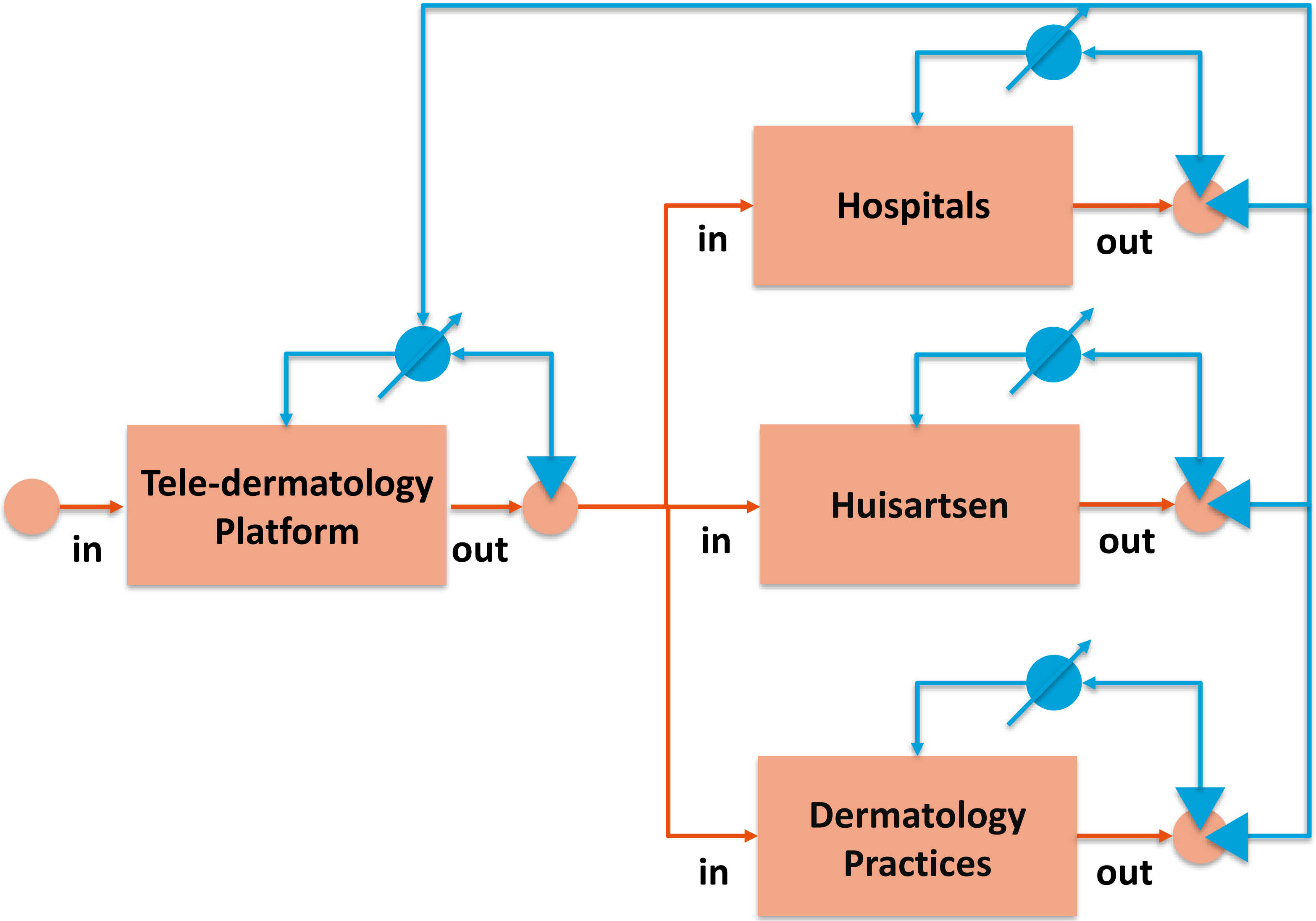


Fig 1: Tele-dermatology service model

Outcome Business



To offer our platform as a service to their patients, enabling remote consultations

To refer patients to dermatology specialists within their network or beyond, improving the referral process

To increase their patient base by offering remote consultations, second opinions, and follow-up services through our platform

Fig 2: Outcome Economy of Tele-dermatology service

Outcome Business

Scope: automated diagnosis of pigmented skin lesions service

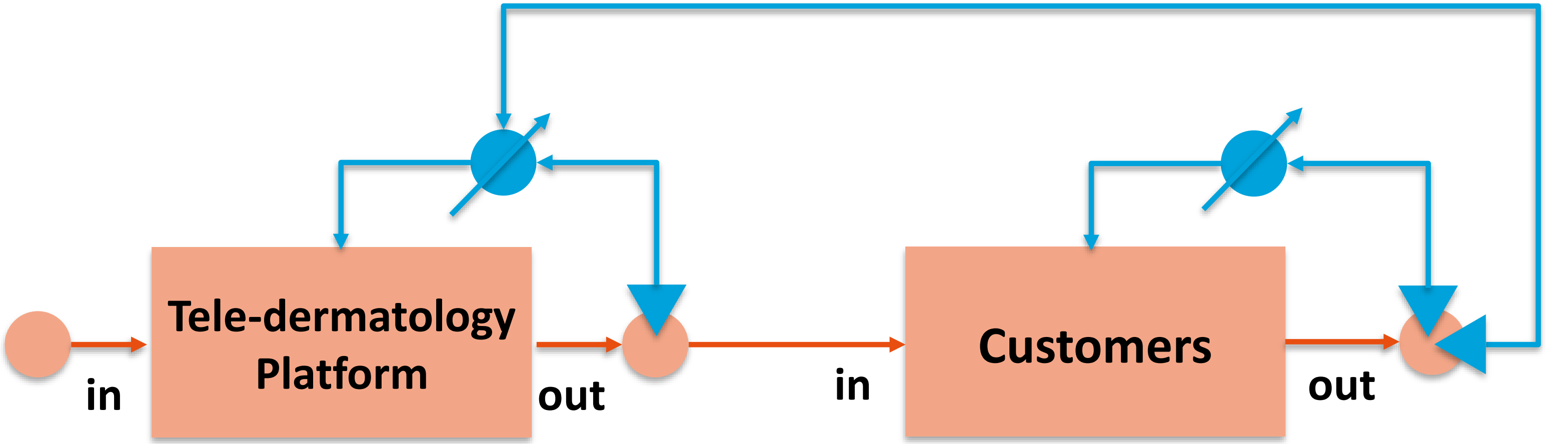


Fig 2: Outcome Economy of Tele-dermatology service

Outcome Business

Scope: automated diagnosis of pigmented skin lesions service

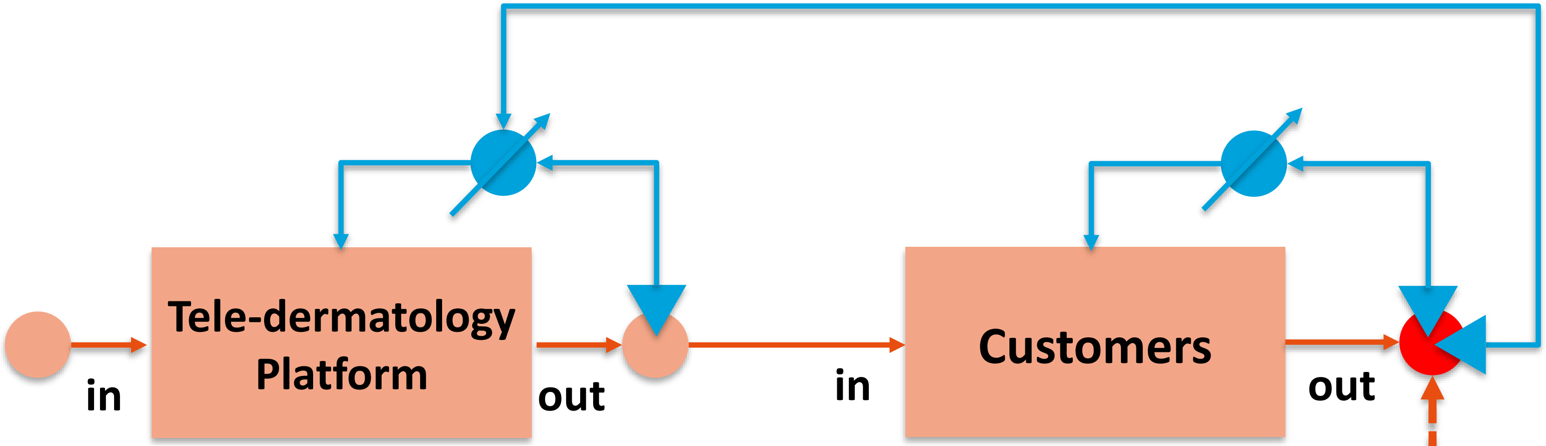


Fig 2: Outcome Economy of Tele-dermatology service

Outcomes

- **Diagnostic quality**
- Image Analysis Efficiency
- Impact on Clinical Workflow
- Patient Care and Outcomes
- System Performance and Reliability
- Economic and Operational Benefits
- User Satisfaction and Acceptance

Outcome Business

Scope: automated diagnosis of pigmented skin lesions service

Product Development

(Strategies to control the provisions that influence our system)

- Enhancing Data Quality
- Model Training and Validation
- Algorithm Optimization
- Performance Benchmarks
- Expert Collaboration
- Ethical and Regulatory Compliance
- Technology and Infrastructure

Quality of Automated Diagnoses

- Evaluate how accurately the system identifies various types of pigmented skin lesions compared to diagnoses made by human experts (Test images).
  Clients' data: 10,015 samples
  Test images: ISIC2018 (1,512 samples)

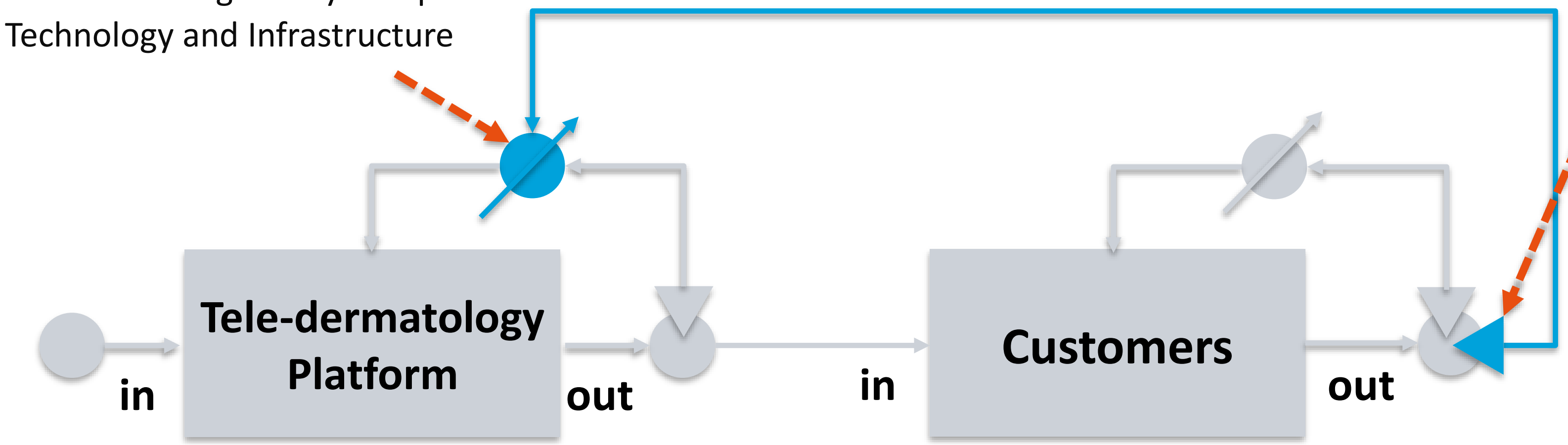


Fig 2: Outcome Economy of Tele-dermatology service

Section 2

Data

Research Question

Type of Data

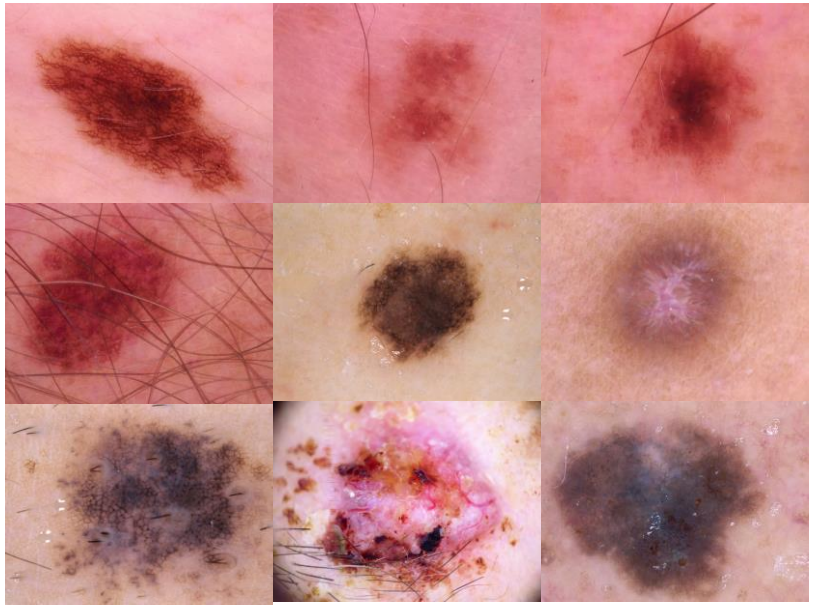


Image Data

- Image ID
- Diagnosis Type (Categorical data)
- Age (Numerical data)
- Sex (Categorical data)
- Localization (Categorical data)
- **Cell type (Categorical data) *Prediction Class**

Tabular Data

Fig 3: Example of the HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions available on https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T

Exploratory Data Analysis

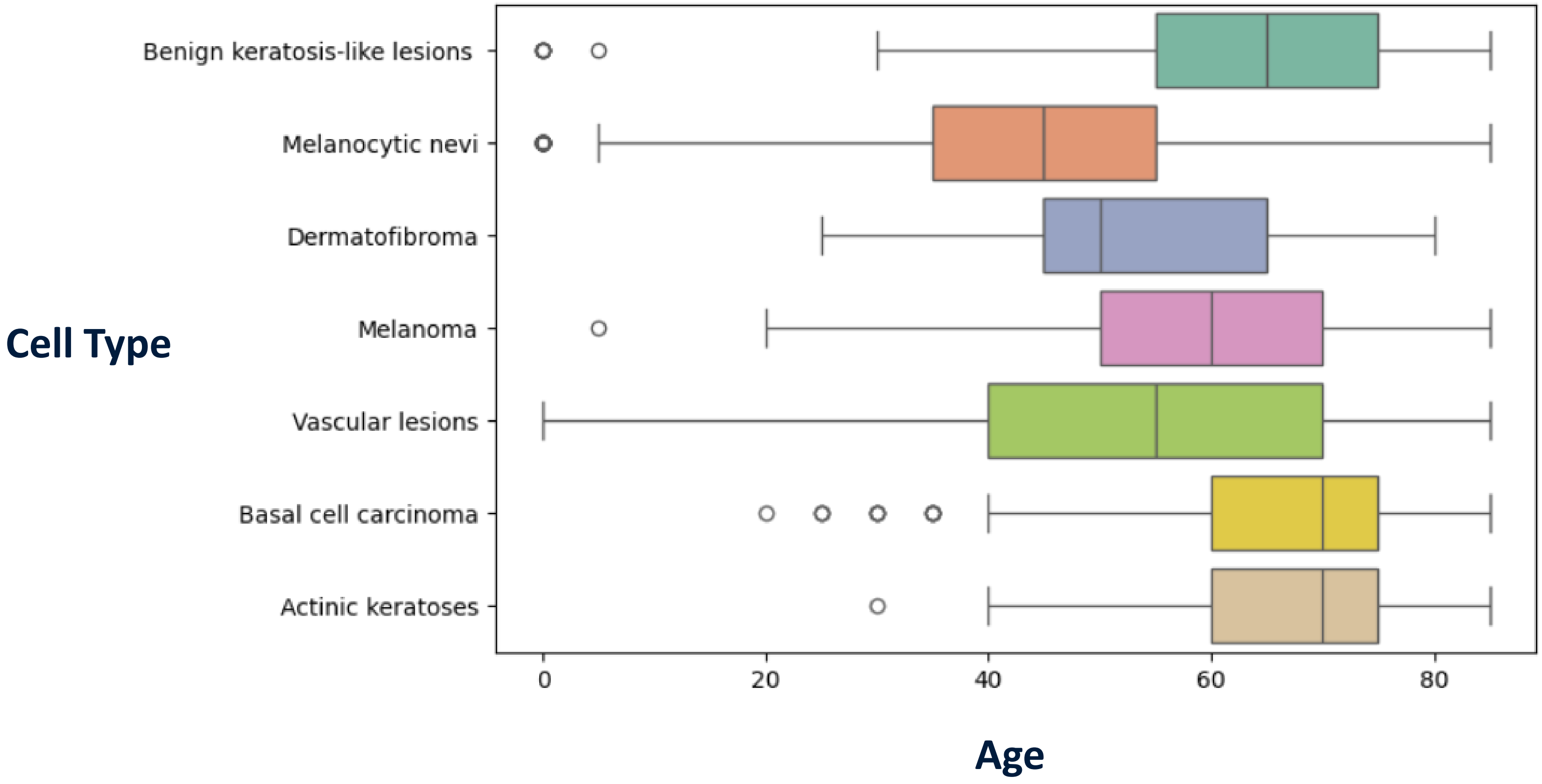


Fig 4: Box Plot of the relationship between “Age” and “Cell Type” Variables

Exploratory Data Analysis (Cramér's V)

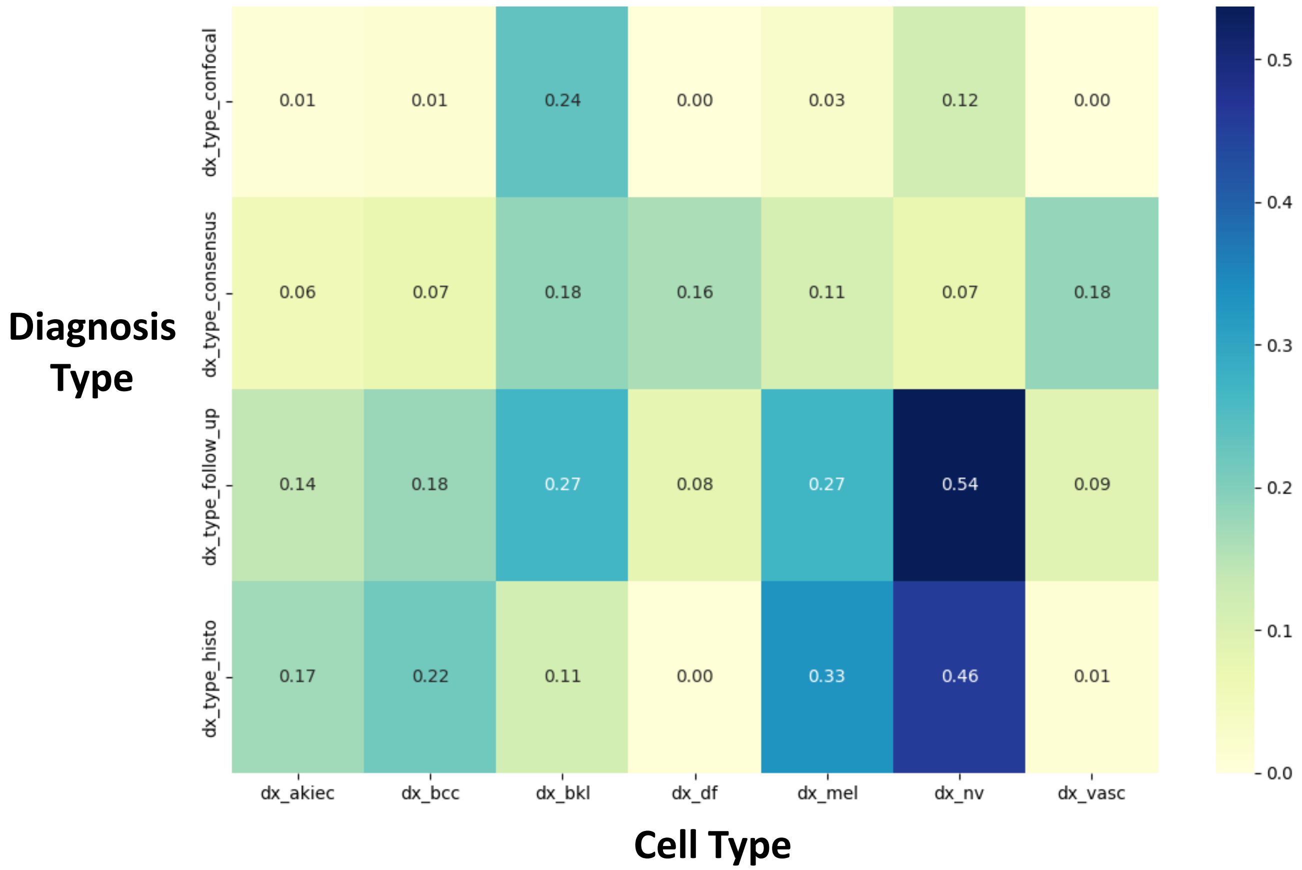


Fig 5: Heatmap of Cramér's V Association between "Diagnosis Type" and "Cell Type" Variables

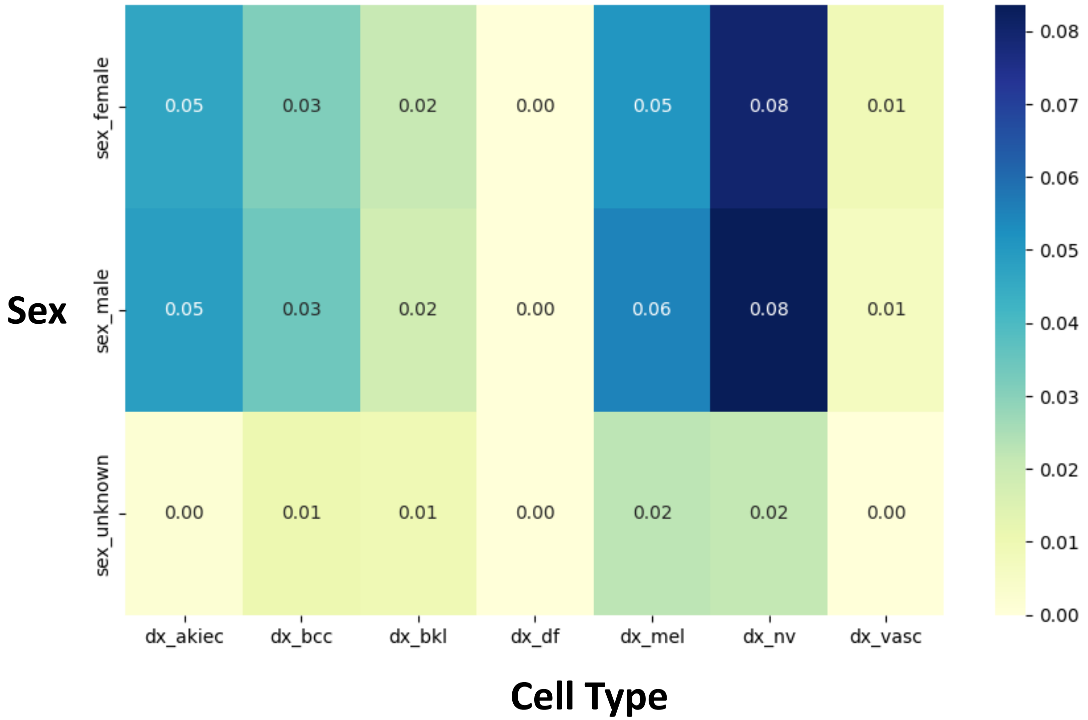


Fig 6: Heatmap of Cramér's V Association between "Sex" and "Cell Type" Variables

Exploratory Data Analysis (Cramér's V)

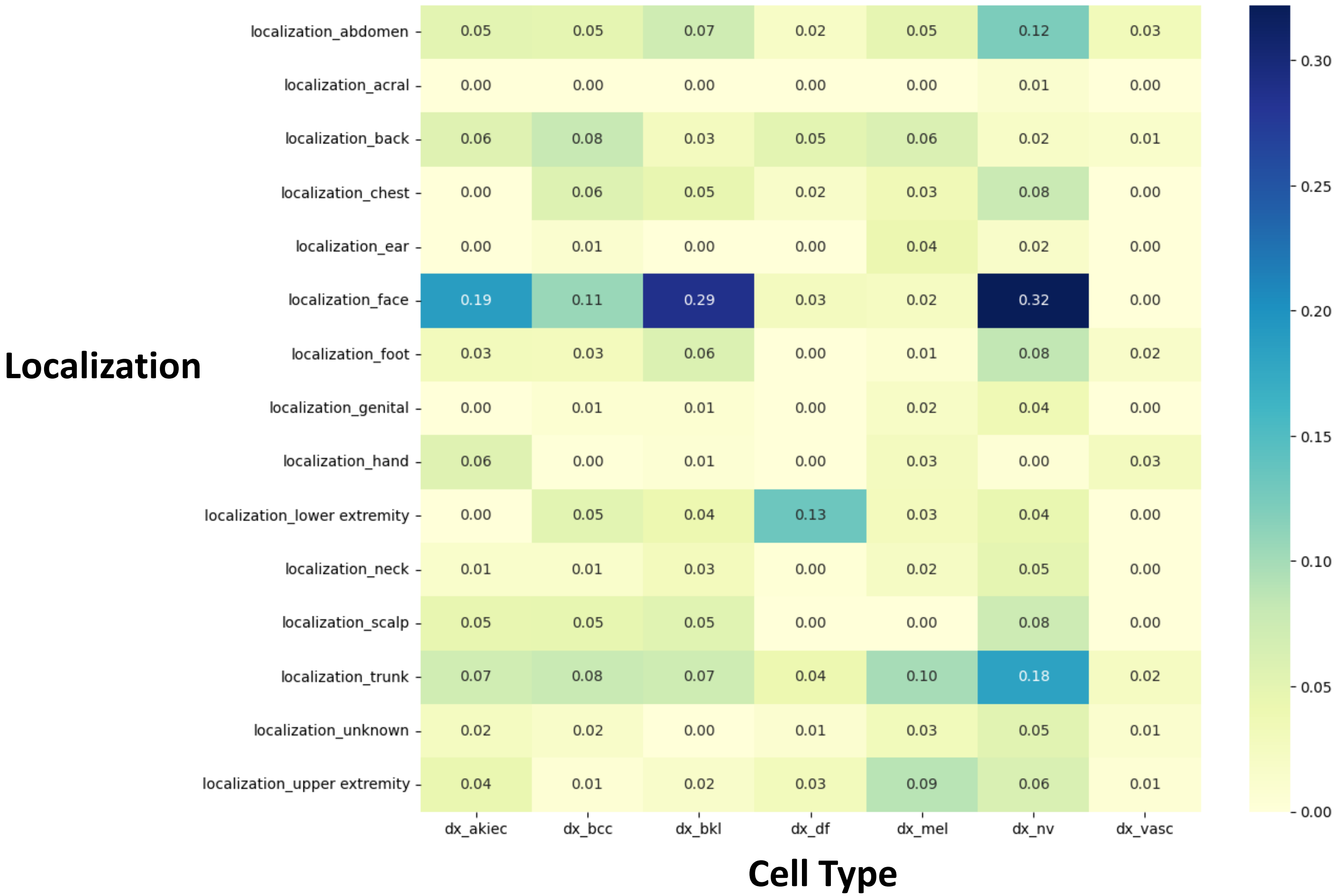


Fig 7: Heatmap of Cramér's V Association between "Localization" and "Cell Type" Variables

We observed correlations between demographic data and cell type.

This suggests integrating demographic data into image classification models might be able to enhance their performance.

Research Questions

- Does fusing “metadata” into a baseline model, initially trained on image features alone, improve its performance ?
- Can Decentralized Model achieve diagnostic accuracy comparable to Centralized Model in the automated diagnosis of pigmented skin lesions?

Why Mid-level Fusion?

- Feature correlation:
From EDA, we find age, diagnosis type, localization in metadata has some correlations with Melanoma skin cancer.

- No low-level fusion:
Difference in datatypes between image and metadata makes it hard to train a single model on the fused raw data.

- No high-level fusion:
Metadata only has four variables, train a model using metadata will not yield high accuracy, then fuse it with the model using image data may not enhance the performance of the latter .

Why Federated Learning?

- Data privacy

Medical data is sensitive, information sharing can be restricted by the regulations.

- Sharing and Storage cost

Image data are large in size, centralized model with data from more hospitals requires large storage space and computational power to process.

- Small dataset for local clients

Local hospitals may not have enough cases to train a classification model, which will lead to overfitting or give non-statistically significant results.

-Non IID dataset for local clients

Data from local hospital may have a certain type of patients, the non-IID features will lead to bias and provide less generalized results.

Points of comparison

Comparative Analysis of Diagnostic Quality

- Quality Metrics
- Lesion Classification Performance

Privacy and Ethical Implications

- Data Privacy
- Trust and Willingness to Share

Impact on Model Generalizability

- Cross-Institution Performance
- Diversity and Representation (Out of Scope *Explainable AI Course)

Section 3

Experimental Setup

Experimental Setup

RQ 1: Does fusing “metadata” into a baseline model, initially trained on image features alone, improve its performance?

Objectives

1. **Evaluate the Impact of Metadata:** Determine if fusing metadata features into an image-based model improves performance metrics.
2. **Identify Key Metadata:** Find out which metadata types most significantly enhance model accuracy.

Baseline	Centralized model (trained on image features only)
Models #1	Centralized model using mid-level fusion (image +metadata)

Table 1: Comparison 1- Baseline Model (Centralized model - trained on image features only) vs. Model #1(Centralized model using mid-level fusion - image and tabular data)

Experimental Setup

RQ 2: Can Decentralized Model achieve diagnostic accuracy comparable to Centralized Model in the automated diagnosis of pigmented skin lesions?

Objectives

Compare Accuracy: Analyze if decentralized models can match centralized models in diagnosing pigmented skin lesions.

Baseline	Centralized model (trained on image features only)
Model #2	Decentralized model using Horizontal Federated Learning (trained on image features only)

Table 2: Comparison 2: Baseline Model (Centralized model - trained on image features only) vs.
Model #2 (Decentralized model using Horizontal Federated Learning - trained on image features only)

Preparing Data for Models #1

Tabular Branch

Tabular Data

Preprocess

- Convert **categorical data** to a **one-hot encoded format**.
- **Normalize numerical data** to have a mean of 0 and a standard deviation of 1.
- Handling **Missing Values**

- **Method 1: Feature Selection**
  Feature Important (Random Forest)
- **Method 2: Feature Extraction**
  Principal Component Analysis (PCA)
- **Method 3: Feature Extraction**
  Artificial Neural Network (ANN)

Image Branch

Image Data
(75 x 100 x 3)

Feature Extraction

MobileNetV2 without the top classification layer

Concatenate Features (Tabular & Image)

Fig 8: Strategy for preparing data for models#1

Experimental Design (Baseline)

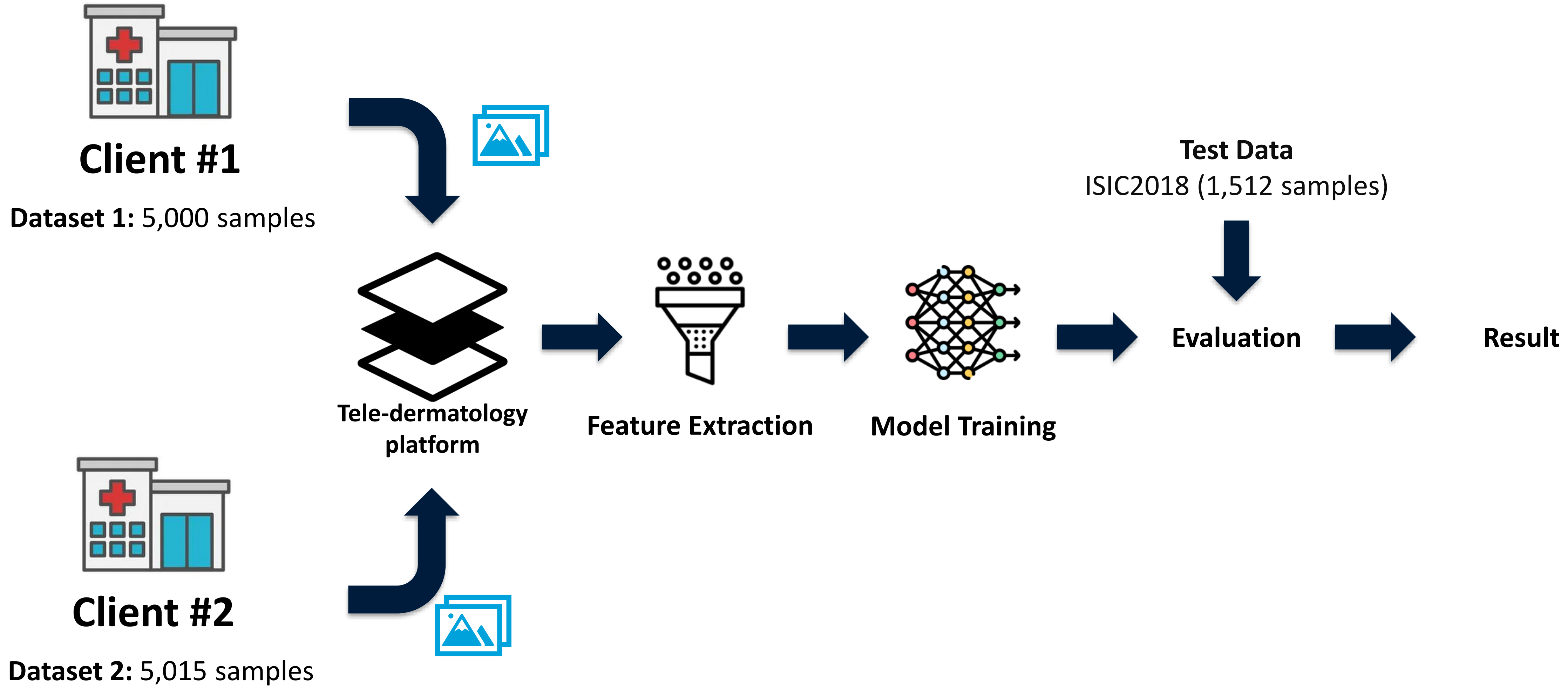


Fig 9: Experimental Design Framework for Baseline & Models #1 Experiments

Experimental Design (Models #1)

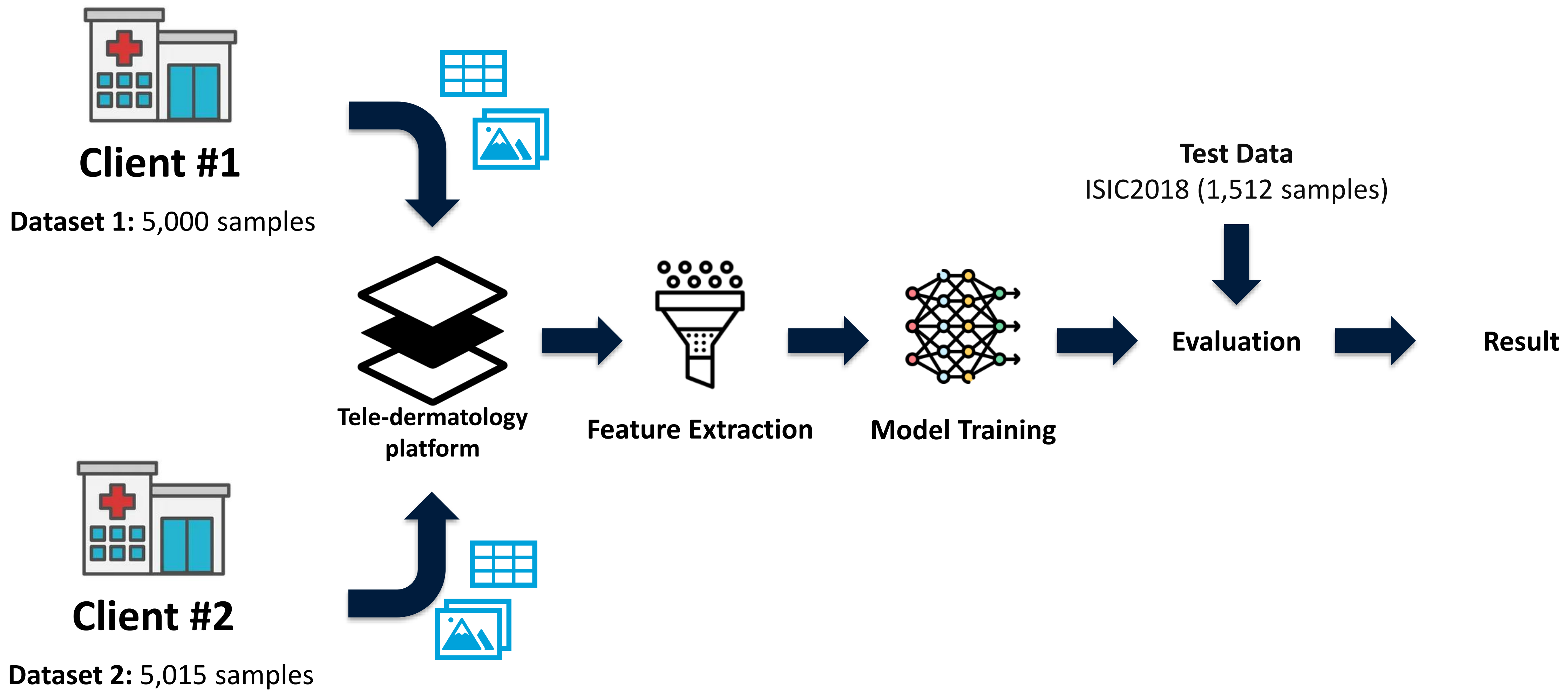


Fig 9: Experimental Design Framework for Baseline & Models #1 Experiments

Experimental Design (Model #2)

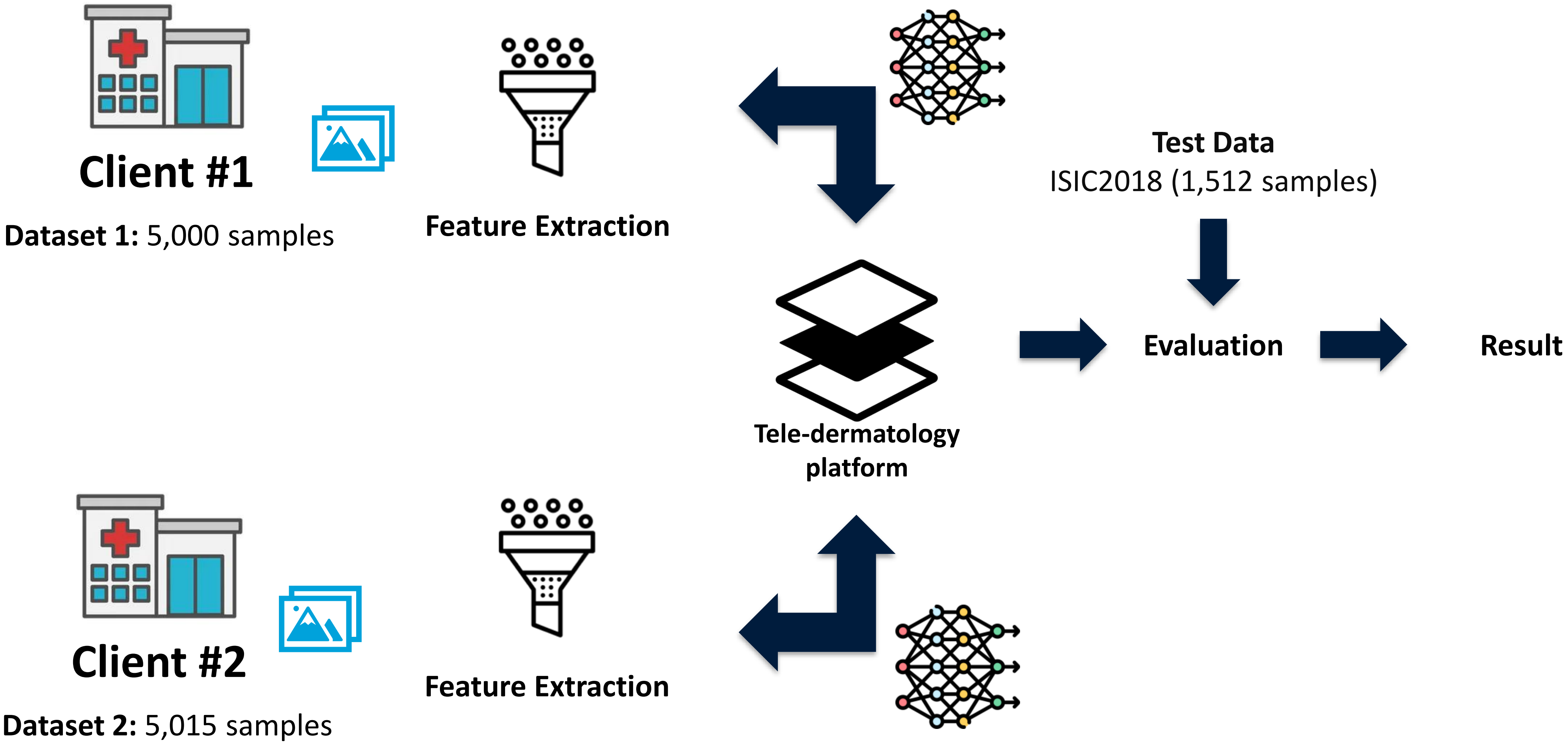


Fig 10: Experimental Design Framework for Model #2 Experiment

Section 4

Methodologies

Image Feature Extraction (MobileNetV2)

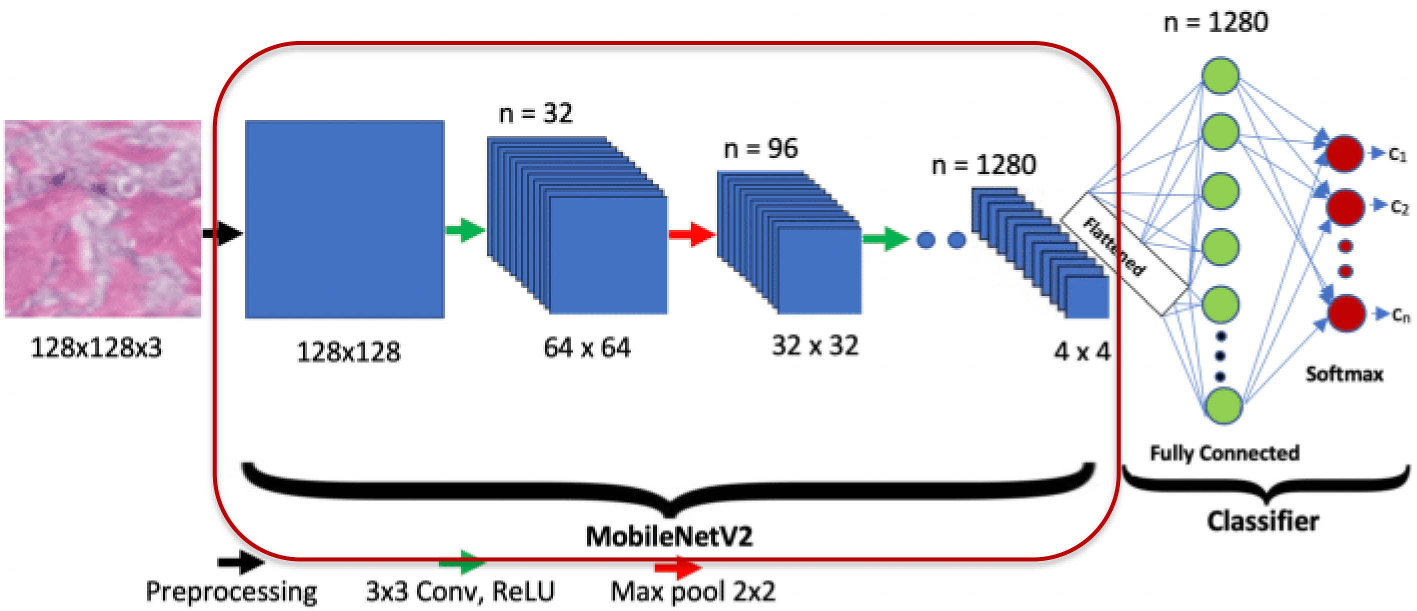


Fig 11: Example of the architecture of the MobileNetV2 (Classification)
https://www.analyticsvidhya.com/blog/2023/12/what-is-mobilenetv2/

- **Depthwise Separable Convolutions**
  - Depthwise Convolution
  - Pointwise Convolution (1x1 Convolution)
- **Inverted Residuals and Linear Bottlenecks**:
  - Bottleneck Layer
  - Inverted Residual Block
- **Feature Extraction**: Sequential layers extract increasingly complex features, with early layers capturing basic shapes and textures, and deeper layers identifying more complex patterns.

SoA: MobileNetV2 (Pretrained: ImageNet)

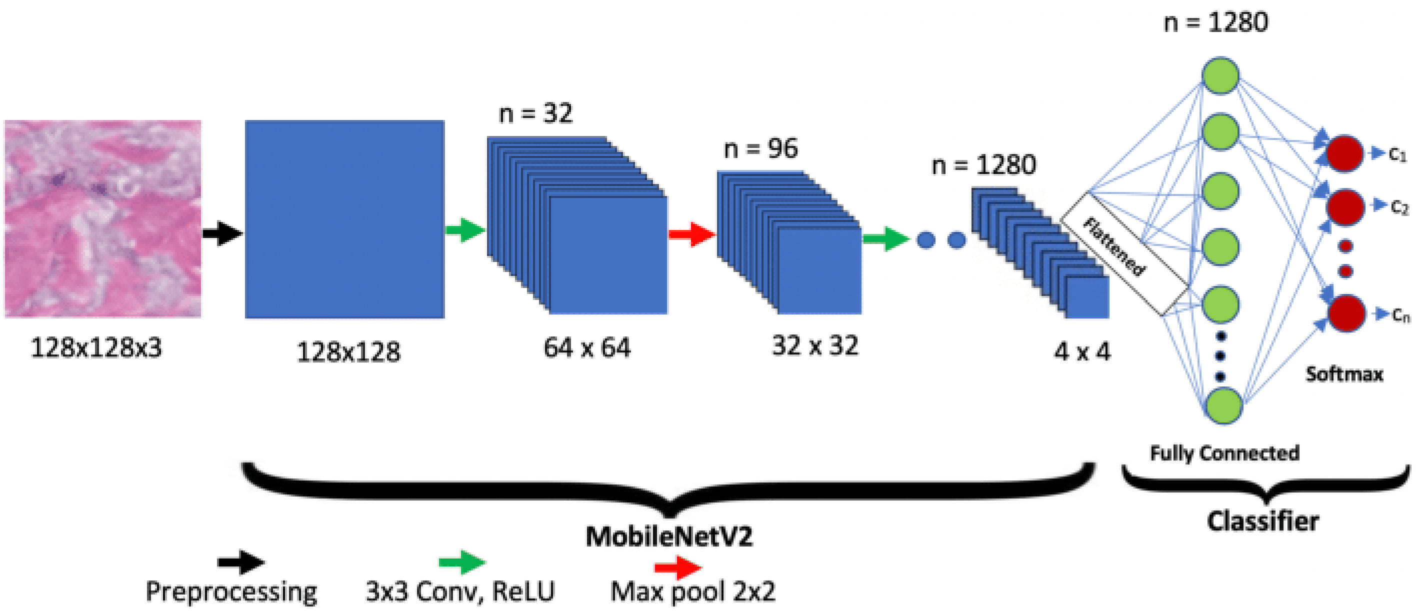


Fig 11: Example of the architecture of the MobileNetV2 (Classification)
https://www.analyticsvidhya.com/blog/2023/12/what-is-mobilenetv2/

- **Output Layer**: Concludes with global average pooling and a fully connected layer, mapping extracted features to the task's output format (e.g., class probabilities).

Tabular Data Feature Selection (Random Forest)

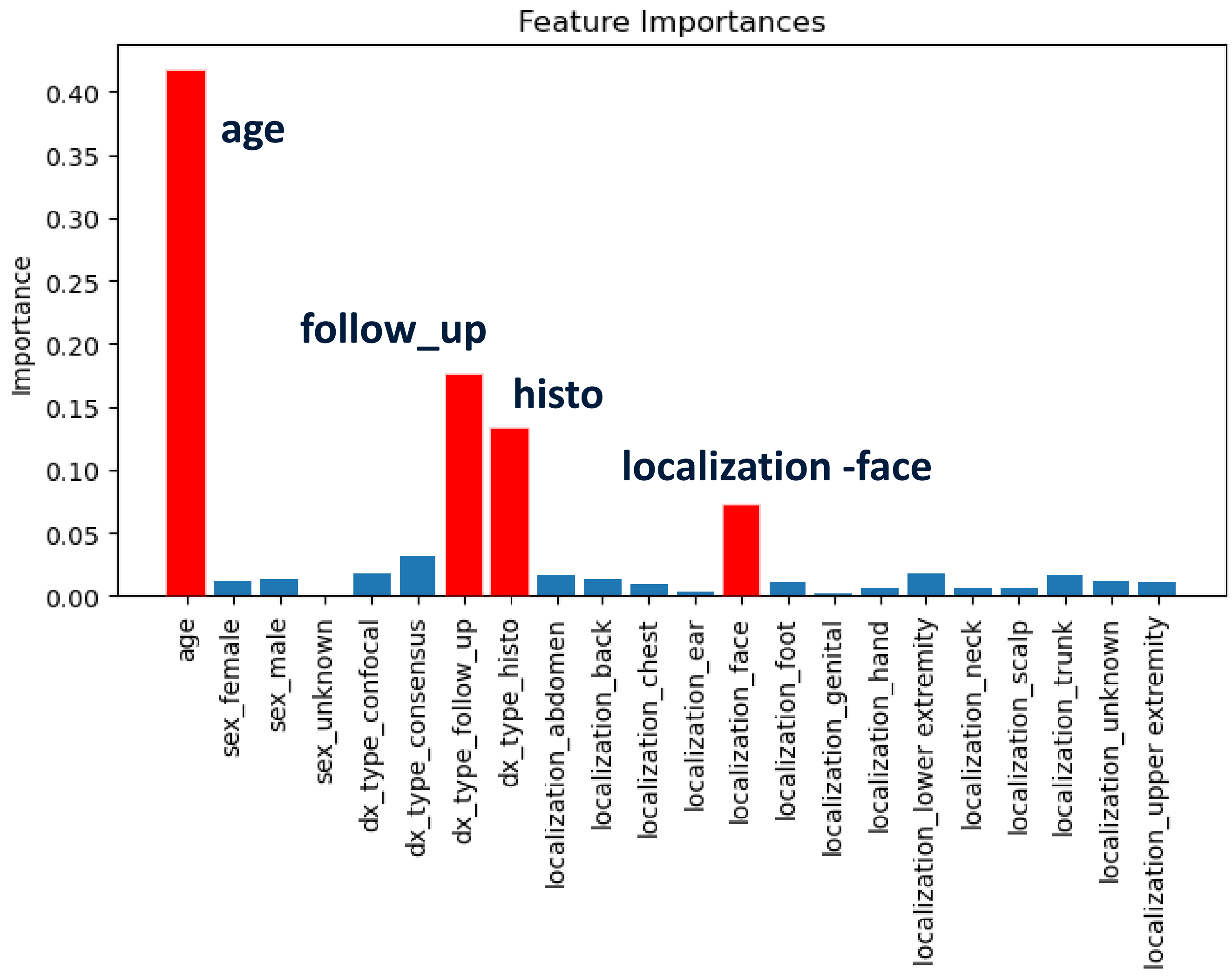


Fig 12: Important features visualization of “Metadata” using Random Forest

Top 4 important features are:

- Age
- dx_type_follow_up
- dx_typr_histo
- localization_face

Tabular Data Feature Extraction (PCA)

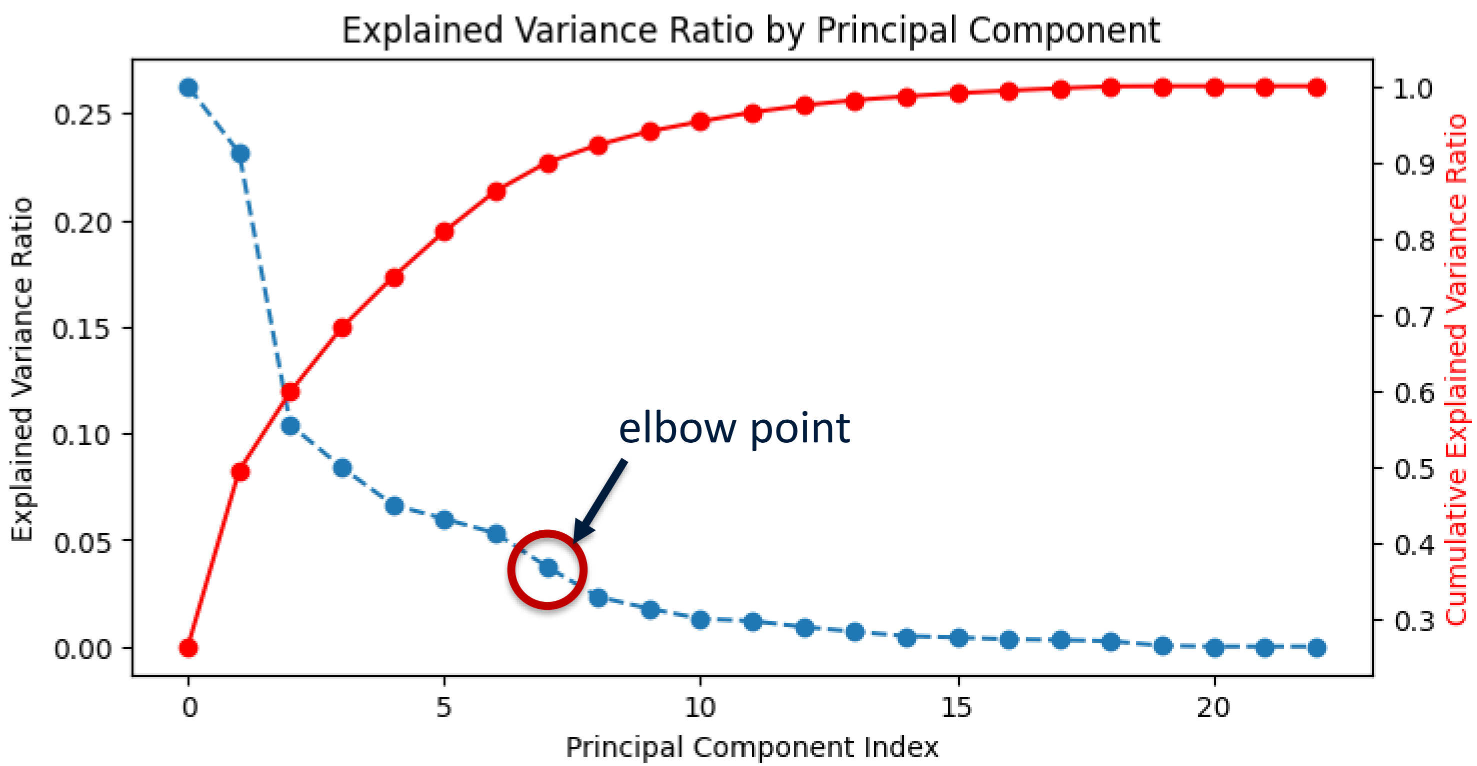
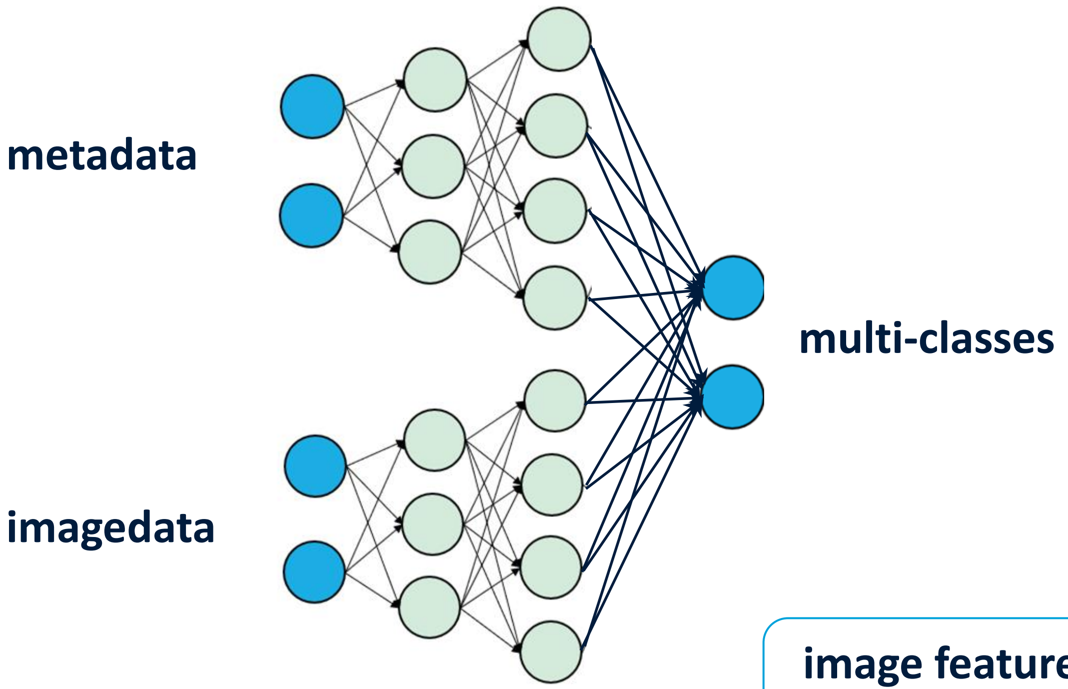


Fig 13: This graph illustrates the Explain Variance Ratio by Principal Component of Tabular Data (Training Data)

The elbow point is at PC8.
So, dimensionality of the data is reduced by keeping the first “8" PCs.

Tabular Data Feature Extraction (ANN)



- Two sequential layers are applied on encoded metadata, generated 512 features.
- those features are then concatenated with 1280 features from image, put into final dense layer for multi-class classification

image features
1280 neurons

✚

metadata 22 neurons ⇄ **1st hidden layer** 1024 neurons ⇄ **2nd hidden layer** 512 neurons ⇄ **Final layer -** sigmoid function 7 classes

Fig 14: Example of Artificial Neural Network (ANN) Implementation

Horizontal Federated Learning

Algorithm 1 FedAvg. K is the total numbers of clients; B is the size of mini-batches, T is the total number of communication rounds, E is the total local training epochs, and η is the learning rate.

Server:
Initialize global model θ_0
for each communication round $t = 1, 2, ...T$ **do**
    Select $m = C \times K$ clients, where $C \in (0, 1)$
    for each **Client** $k = 1, 2, ...m$ in parallel **do**
        Download θ_t to **Client** k
        Do **Client** k update and receive θ^k
    end for
    Update global model $\theta_t \leftarrow \sum_{k=1}^{m} \frac{n_k}{n} \theta^k$
end for

Client k update:
Replace local model $\theta^k \leftarrow \theta_t$
for local epoch from 1 to E **do**
    for batch $b \in (1, B)$ **do**
        $\theta^k \leftarrow \theta^k - \eta \nabla L_k(\theta^k, b)$
    end for
end for
Return θ^k

Fig 15: Algorithm of Horizontal Federated Learning (FedAvg)

Section 5

Results

Discussion (Conclusion)

- Comparative Analysis of Diagnostic Quality
- Privacy and Ethical Implications
- Impact on Model Generalizability
- Technological and Methodological Innovations

Discussion: Comparative Analysis of Diagnostic Quality

Baseline vs. Models #1 (Centralized Model)

	Input	Cell Type Prediction (Weighted Average)			Cell Type Prediction - Melanoma Prediction		
		Precision	Recall	F1-score	Precision	Recall	F1-score
Baseline	**Image Only**	**0.69**	**0.47**	**0.49**	**0.57**	**0.30**	**0.39**
Mid-level Fusion (Models #1)	Image Feature + Metadata	0.65	0.39	0.40	0.51	0.20	0.29
	Image Feature + Metadata Feature(RF)	**0.62**	**0.52**	**0.55**	0.55	0.36	0.43
	Image Feature + Metadata Feature(PCA)	0.65	0.49	0.52	**0.64**	**0.36**	**0.46**
	Image Feature + Metadata Feature(ANN)	**0.65**	**0.53**	**0.57**	**0.51**	**0.51**	**0.51**

Table 3: Results Comparison of model quality metrics between Baseline Model (Centralized model - trained on image features only) vs. Model #1(Centralized model using mid-level fusion - image and tabular data)

Discussion: Comparative Analysis of Diagnostic Quality

Baseline vs. Model #2 (Decentralized Model)

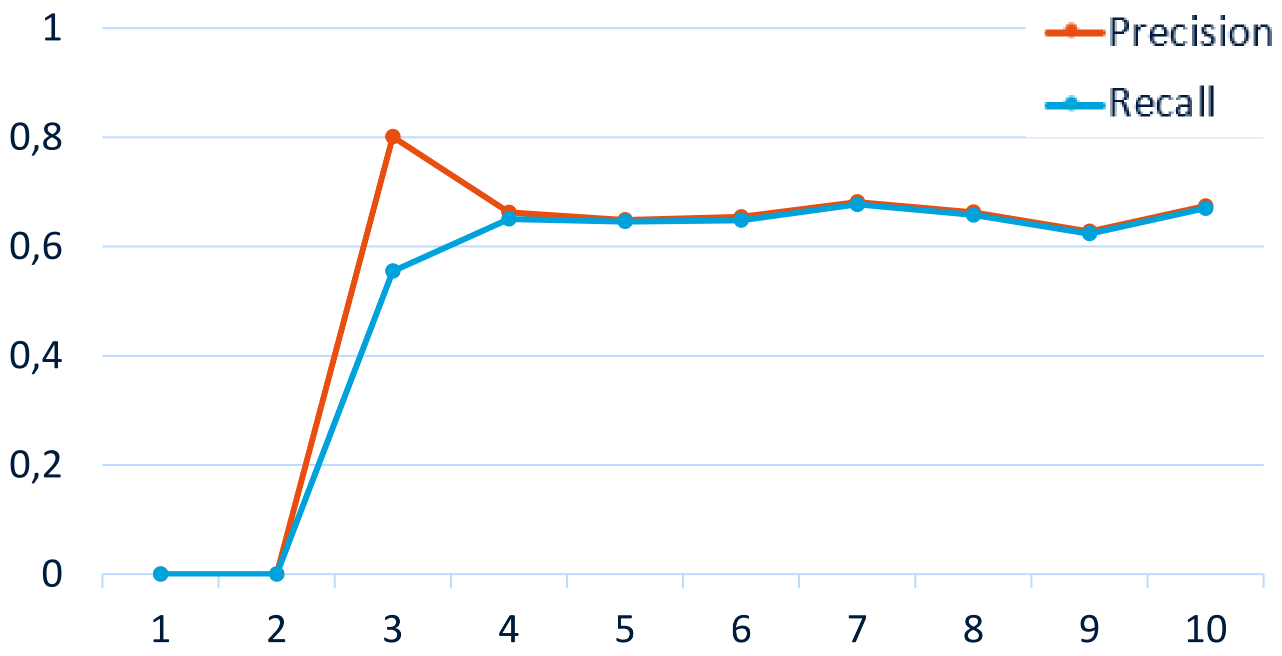


Fig 16: Federated Model Performance for 10 communication rounds

	Cell Type Prediction (Weighted Average)		
	Precision	Recall	F1-score
Baseline	0.69	0.47	0.49

Table 4: Baseline model performance results

Model Performance Metrics

- Precision and Recall improve over communication rounds, showing learning progress.
- By the final round, Recall (0.67) and F1-Score (0.67) values both exceed the metrics of Baseline. *Precision is 0.67

Discussion

Baseline vs. Model #2 (Decentralized Model)

Learning Efficiency: Convergence Rate

- Fig 16 suggests a non-linear improvement in model performance metrics over communication rounds
- This can be contrasted with the centralized model, which doesn't benefit from iterative improvements post-training unless retrained

Data Privacy and Distribution

- Federated learning (Model #2) has a significant advantage in terms of privacy
- Fig 16 shows that even with distributed data, effective learning can occur.

Scalability and Real-World Applicability

- Baseline relies on centralized data collection, which may not always be feasible or ethical
- Model #2 reflects a more scalable approach in scenarios where data cannot be centralized due to privacy concerns, bandwidth limitations, or regulatory restrictions.

Section 6

Research Gaps
Future Research

Research Gaps & Future Research

Data Privacy and Security in Federated Learning

- Developing more advanced encryption and secure aggregation techniques to enhance privacy without significantly impacting model performance or training efficiency.

Integration of Heterogeneous Data Types

- Exploring more sophisticated feature fusion techniques that can handle heterogeneity more effectively, ensuring that the integration of different data types leads to actual performance gains.

Efficiency and Scalability of Federated Learning

- Investigating methods to reduce communication costs, improve model convergence rates, and efficiently scale federated learning systems across many nodes.

Bias and Fairness

- Developing methodologies for detecting, quantifying, and mitigating bias in machine learning models.

Research Gaps & Future Research

Generalization Across Diverse Datasets

- Enhancing model robustness and generalization capabilities, possibly through more diverse training data, transfer learning techniques, or advanced regularization methods.

Optimal Model and Hyperparameter Selection

- Automating the process of model selection and hyperparameter tuning, possibly through meta-learning or AutoML techniques, to optimize performance across various tasks and data distributions.

Interpretability and Explainability

- Developing tools and techniques for model interpretability and explainability, ensuring that users can understand and trust the decisions made by machine learning models, especially in critical applications.

References

1. Flower Labs GmbH. (n.d.). *Flower: A Friendly Federated Learning Framework*. Retrieved February 5, 2024, from https://flower.ai/
2. Holste, G., Partridge, S. C., Rahbar, H., Biswas, D., Lee, C. I., & Alessio, A. M. (2021). End-to-end learning of fused image and non-image features for improved breast cancer classification from mri. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3294-3303).
3. McMahan, B., Moore, E., Ramage, D., Hampson, S. & Arcas, B.A.y.. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, *in Proceedings of Machine Learning Research* 54:1273-1282 Available from https://proceedings.mlr.press/v54/mcmahan17a.html.
4. Tschandl, P., Rosendahl, C. & Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data 5, 180161 (2018). https://doi.org/10.1038/sdata.2018.161

Appendix

Model Hyperparameters

	Network structure	optimizer	loss fuction	input size	output classes	learning rate	training epochs	batch size	communication rounds
Baseline	MobileNetV2	Adam	binary cross entropy loss	(75,100,3)	7	0.0001	10	32	-
Mid-level Fusion (PCA)	MobileNetV2	Adam	binary cross entropy loss	(75,100,3)	7	0.0001	10	32	-
Mid-level Fusion (ANN)	MobileNetV2 + ANN	Adam	binary cross entropy loss	(75,100,3)	7	0.0001	10	32	-
Federated Learning	MobileNetV2	Adam	binary cross entropy loss	(75,100,3)	7	0.0001	10	32	10

Table 5: Model Hyperparameters of each experiment