



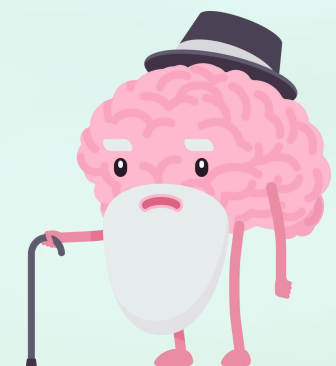
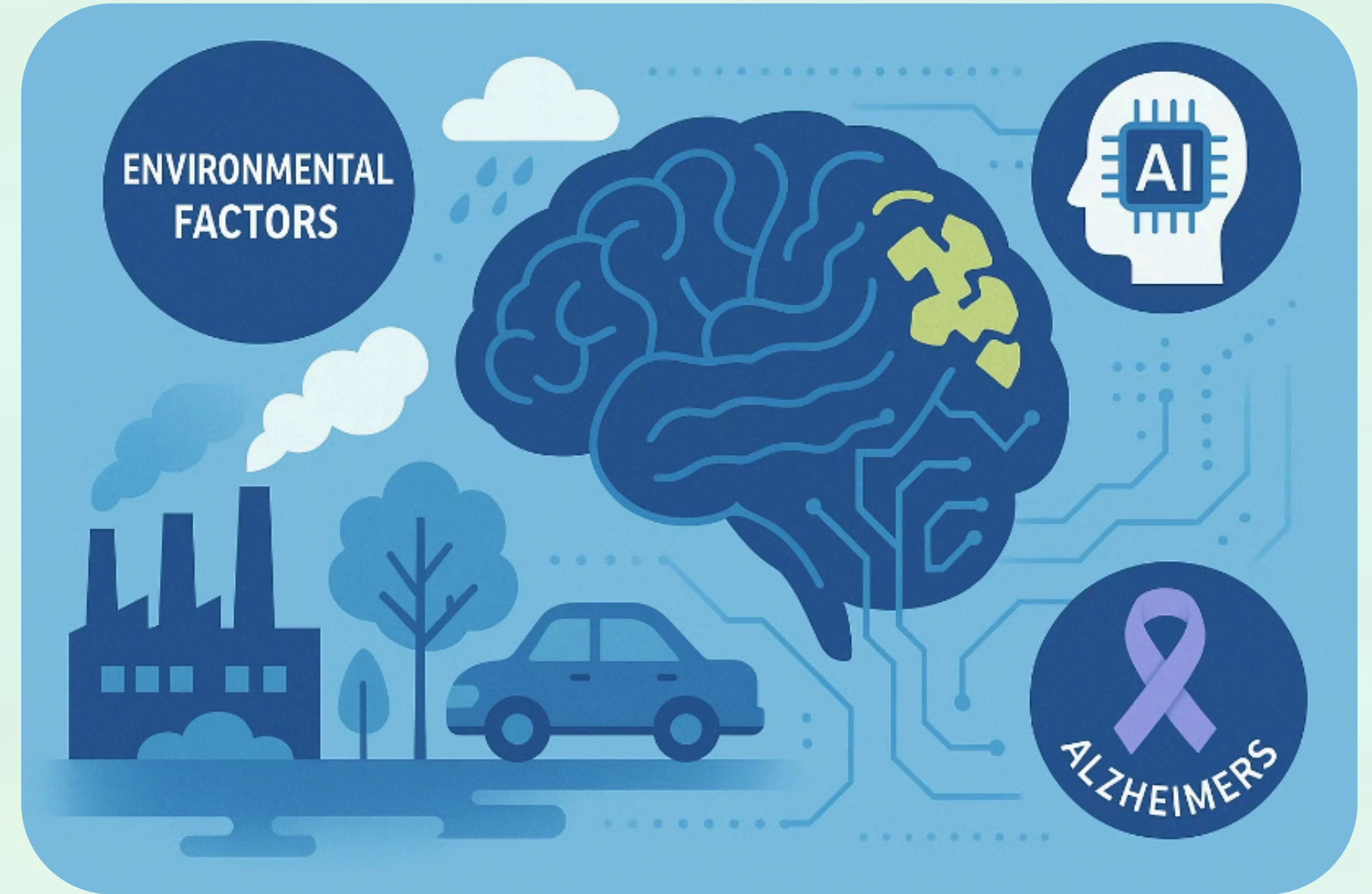
When the Environment Meets the Brain: Uncovering Alzheimer's Risk

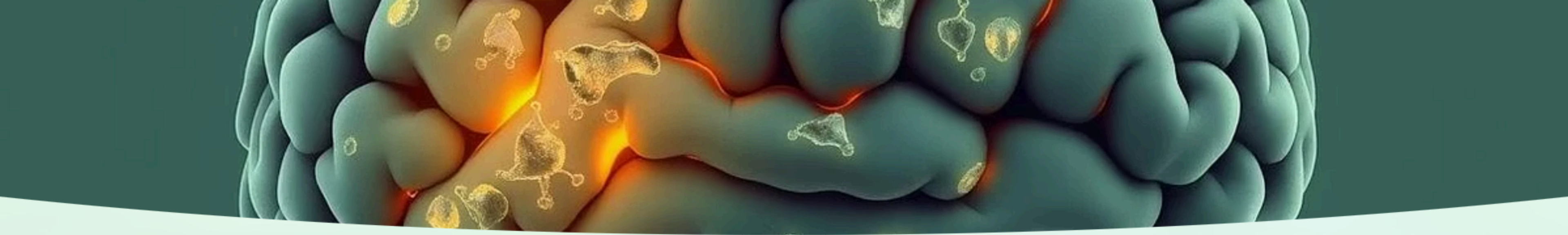
Students:

Amit Eliya,
Amit Kriaf,
Tzlil Septon

Advisors:

Prof. Adi Shraibman,
Prof. Gideon Dror,
Dr. Dorit Shweiki,
Dr. Yonatan Bilu





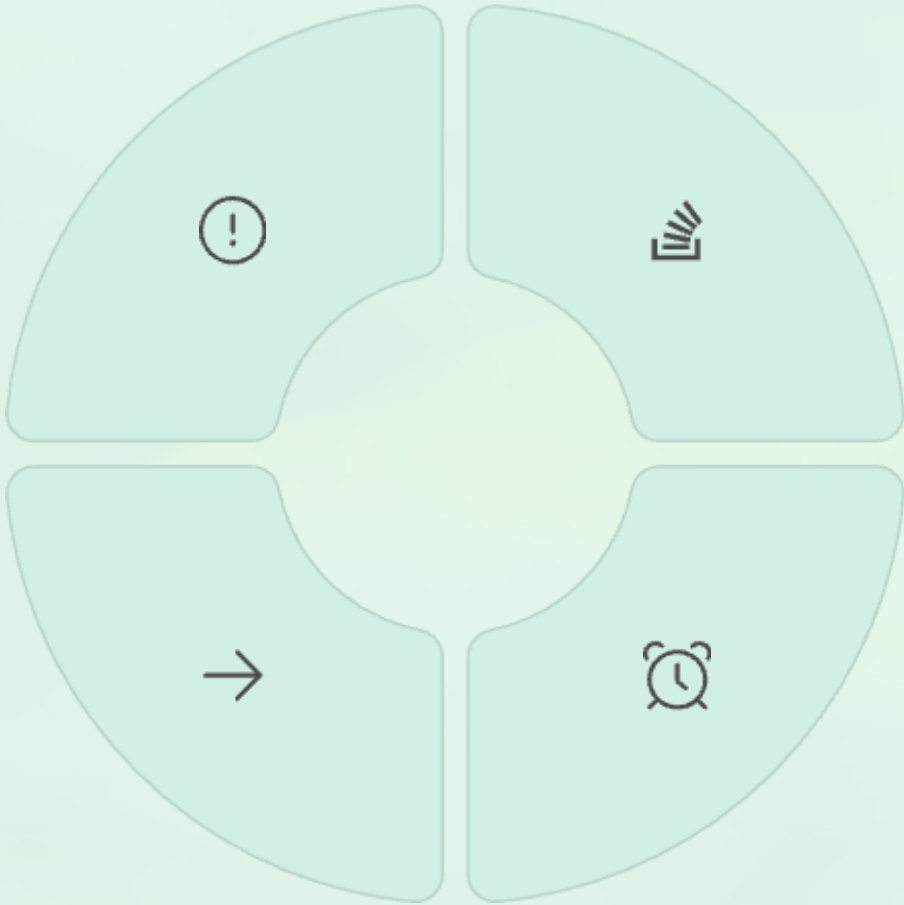
Introduction to Alzheimer's Disease

Various factors

Genetic, lifestyle, and environmental factors all contribute to AD development.

Development of AD

The disease develops gradually, and by the time symptoms appear, significant brain damage may have already occurred. Therefore, early detection and diagnosis are critical for improving patient outcomes and planning care.



Progressive Neurodegeneration

AD is a progressive neurodegenerative disorder that primarily affects memory, thinking, and behavior, eventually impairing the ability to carry out daily tasks.

Early Intervention

The purpose of this study is the early identification of individuals at risk of developing Alzheimer's disease, in order to enable timely intervention and potentially slow disease progression.



UK Biobank: A Resource for Research

Comprehensive Biomedical Database

Established in the UK with over 500,000 participants aged 40-60 at recruitment.

Extensive Health Assessments

Physical measurements and biological samples collected for in-depth analysis.

Detailed Participant Data

Lifestyle information, medical history, and socio-demographic data via questionnaires.

Comprehensive Data Fields for AD Risk

Physiological
Sex
Year of date
Skin color

Environmental
Living environment score (England).

Demographical
Ethnic background.

Health Data
Dementia diagnosis dates
Heart attack age
Memory scores.

Behavioral Factors
Sleep patterns (sleeping during daytime..)
Mobile phone use length.
Fluid intelligence score.
Household size.
Dietary changes and coffee intake.
Time spent using computer or watching TV.
Physical activity duration: moderate, vigorous, walks.
Lifetime number of sexual partners.
Time spent outdoors in summer.
Qualifications.
Use of sun/UV protection.
Smoking status: ever smoked.
Alcohol intake frequency.

XGBoost Model for Predicting Alzheimer's Risk



Decision Trees

The algorithm builds on powerful decision trees, enhancing interpretability.



Boosting Technique

Sequential learning corrects prior tree errors, improving overall robustness.



High Accuracy

XGBoost achieves strong predictive performance, crucial for medical diagnostics.



Complexity Control

Fine-tuned parameters help prevent overfitting, ensuring generalization.

XGBoost (Extreme Gradient Boosting) is an advanced machine learning algorithm that combines accuracy, flexibility, and efficiency. These strengths make it particularly well suited for medical data, where diverse features must be analyzed to produce reliable predictions that support early detection and improved patient outcomes.



Preparing The Dataset for Analysis



Identify Alzheimer's Cases

Filtered UK Biobank data to pinpoint diagnosed Alzheimer's cases. This creates the positive class for our model.



Focus on Late-Onset AD

Included only participants aged 70+ for late-onset AD, the most common form. This ensures relevant patient cohort focus.



Utilize Control Group

Used nearly 500,000 non-Alzheimer's participants as controls. This simulates real-world prevalence of the disease.

Thorough dataset preparation ensures the model captures real-world patterns effectively.

Focusing on late-onset Alzheimer's disease and including a large control group improves the study's external validity and applicability.

Strategic Feature Selection



Relevance Variables

Selected environmental and lifestyle variables pertinent to AD risk.



Data Quality Filtering

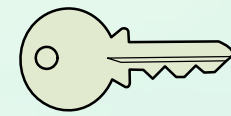
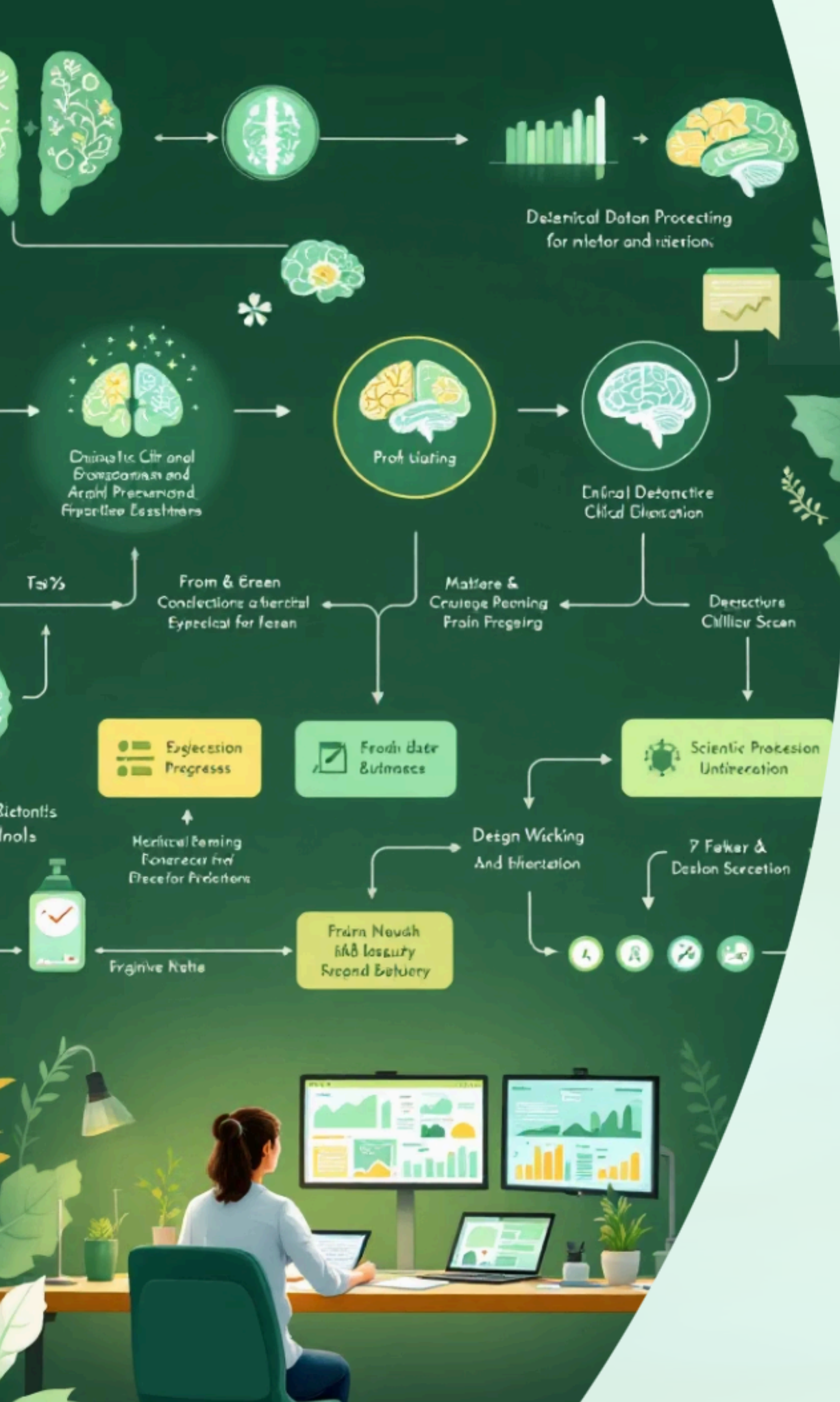
Features with too many missing values or low data quality were excluded.



Data Quality Filtering

From the remaining variables, we kept only the most informative ones to optimize model performance and reduce noise.

This meticulous feature selection process was crucial for building an effective predictive model. It helps to reduce noise and ensures that the model learns from the most impactful variables. This leads to more robust and accurate risk assessments.



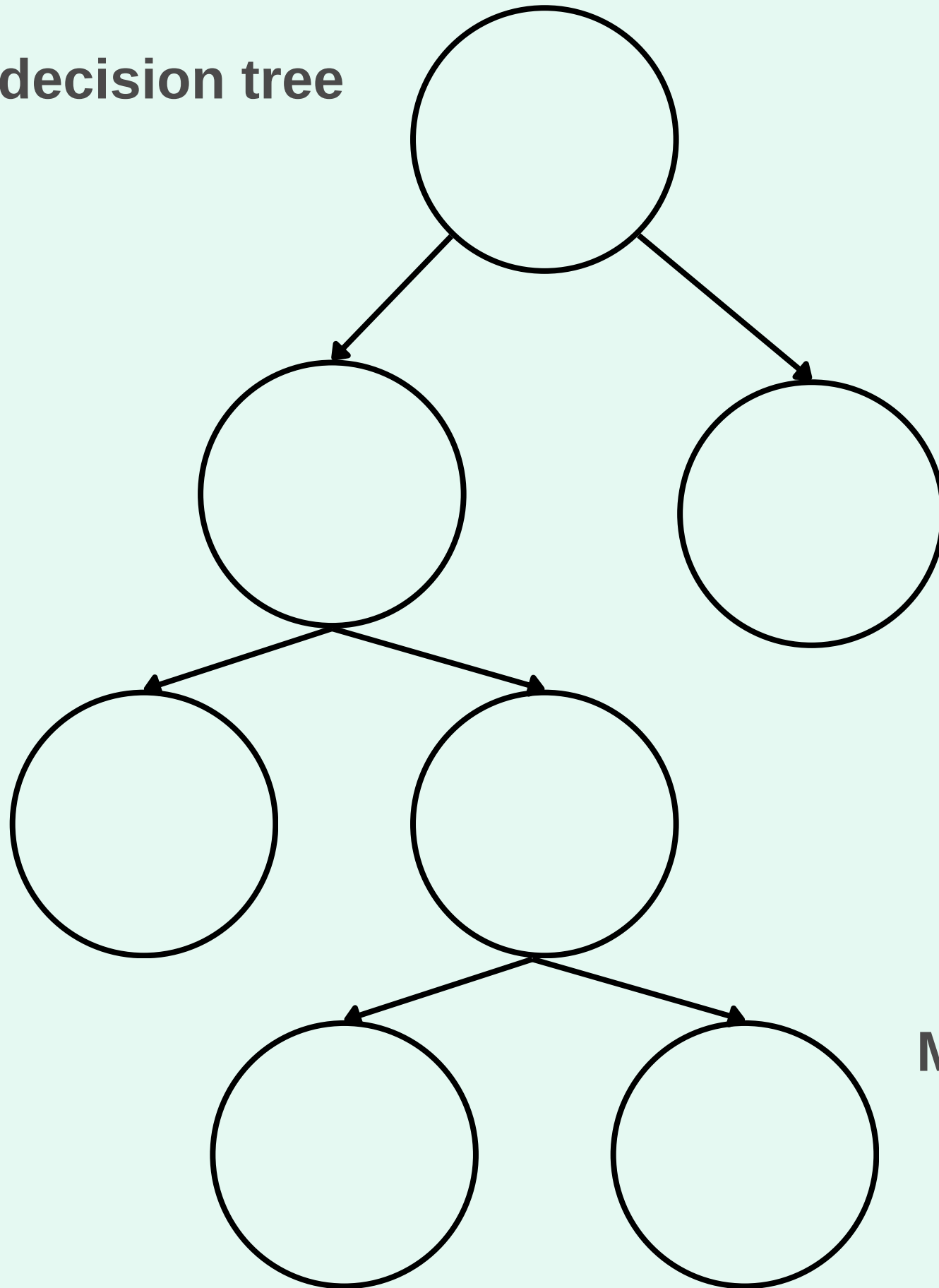
Key Model Parameters

Parameter	Description	Role
n_estimators	Number of trees	More trees improve accuracy but slow down training
max_depth	Maximum depth of each tree	Taller trees learn complex patterns but risk overfitting
learning_rate	Step size for each tree	Smaller values lead to slower, more careful learning.
subsample	% of data used for each tree	Uses part of data to add randomness
colsample_bytree	% of features used per tree	Chooses a random subset of features for each tree.

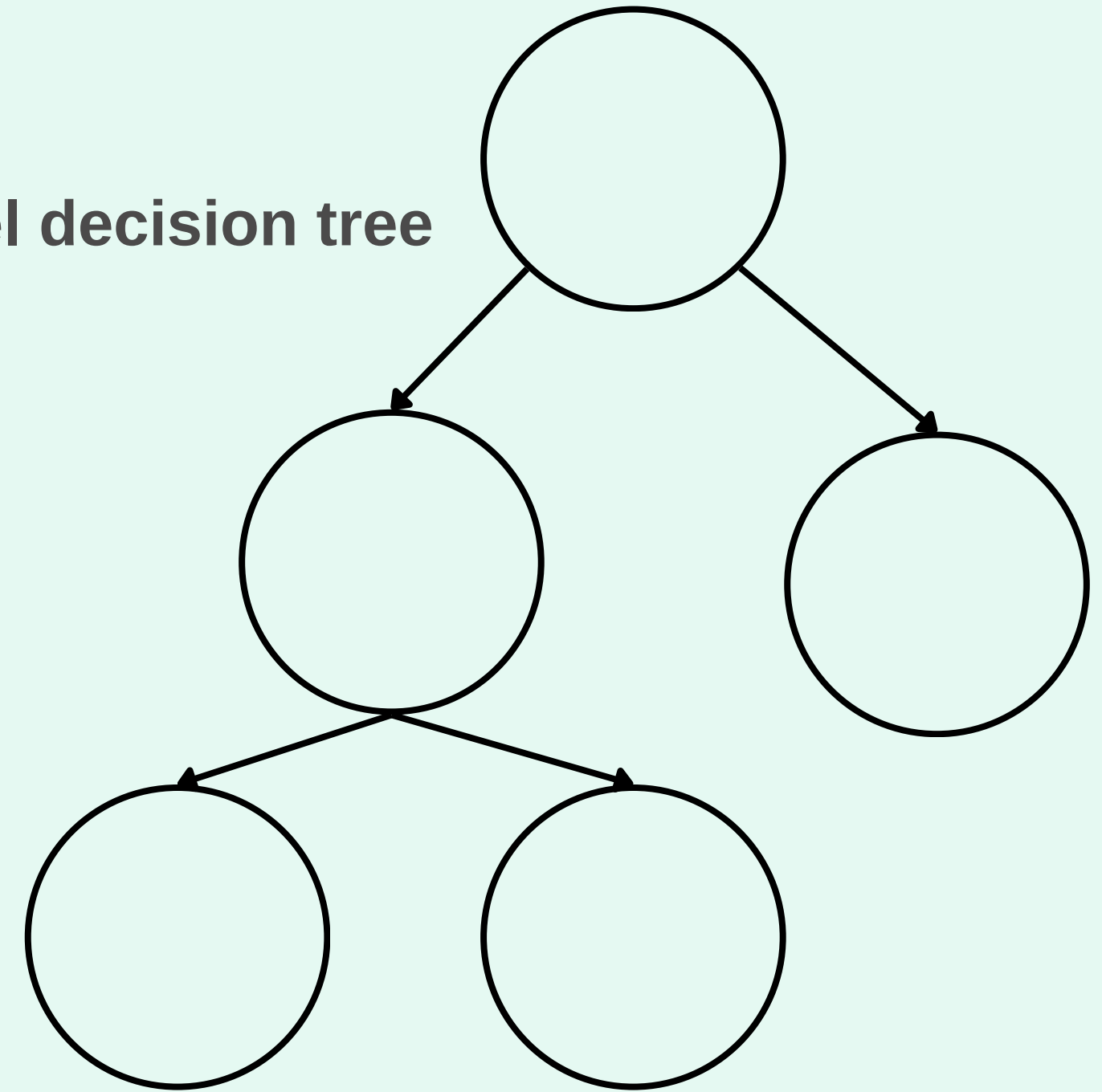
Choosing parameters carefully is key to building a successful model.

Why Tree Height Matters

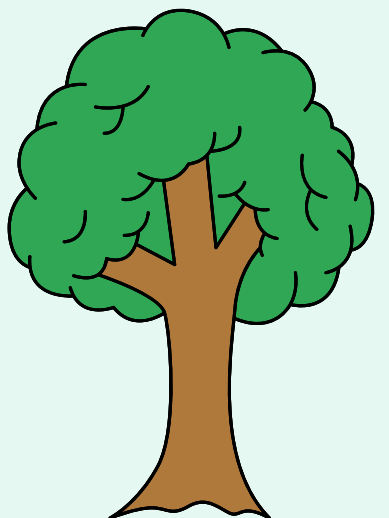
3-level decision tree



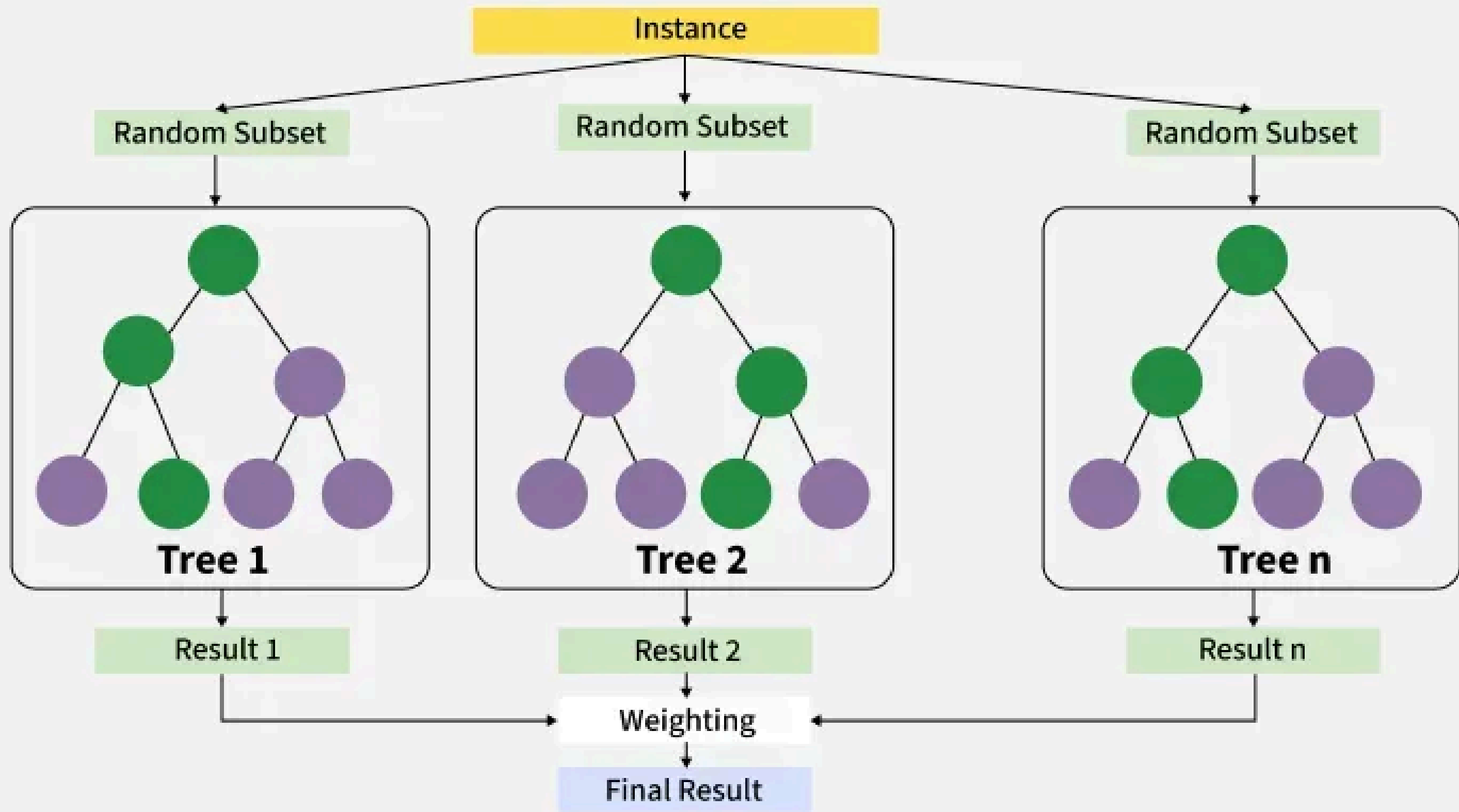
2-level decision tree



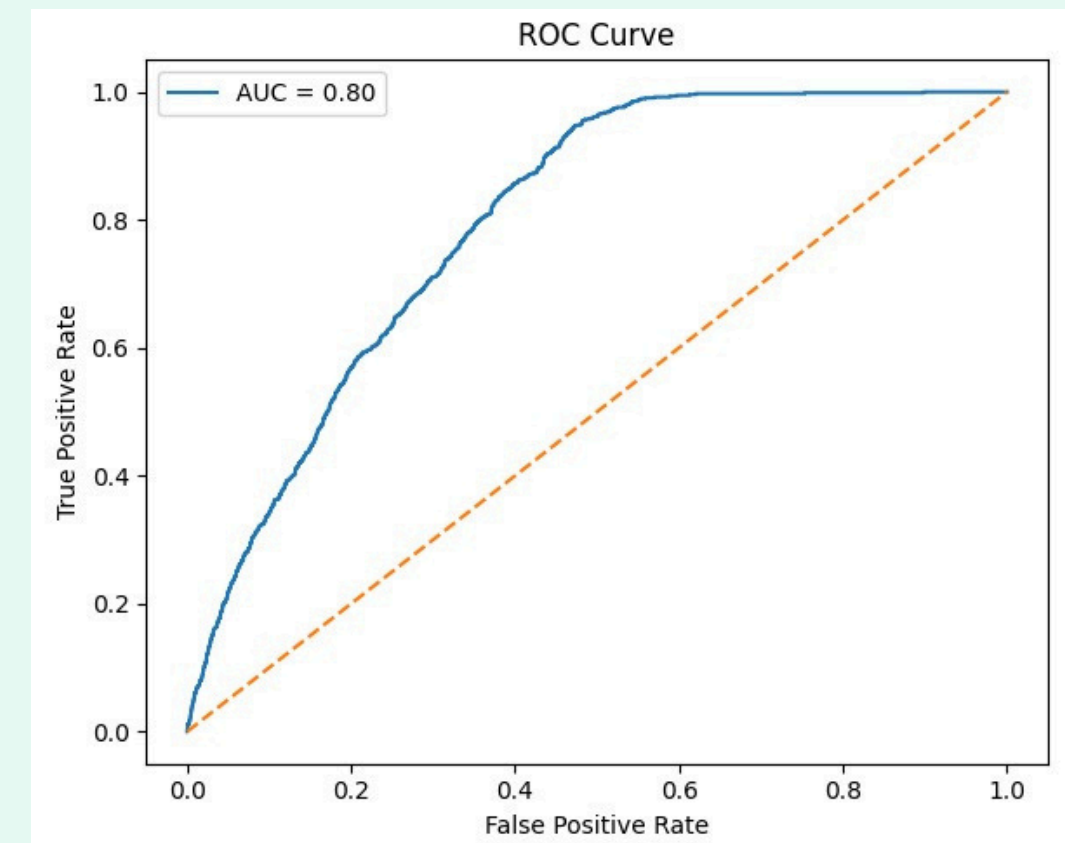
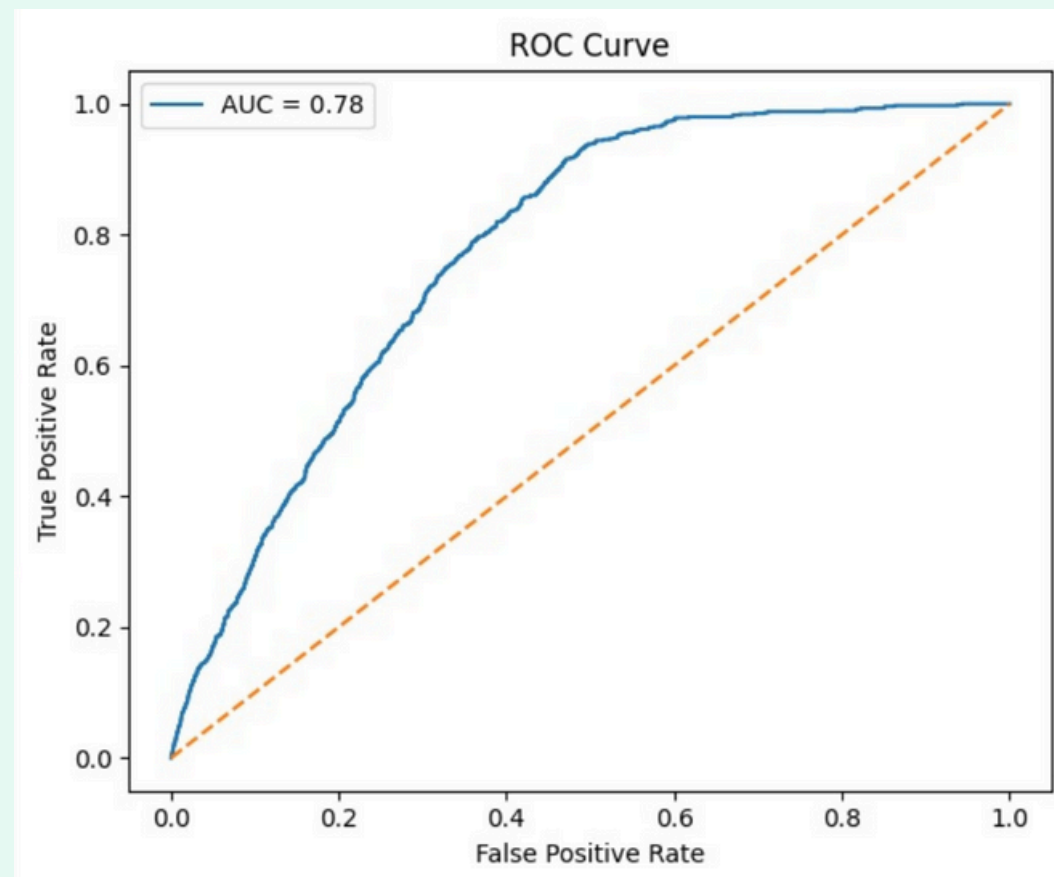
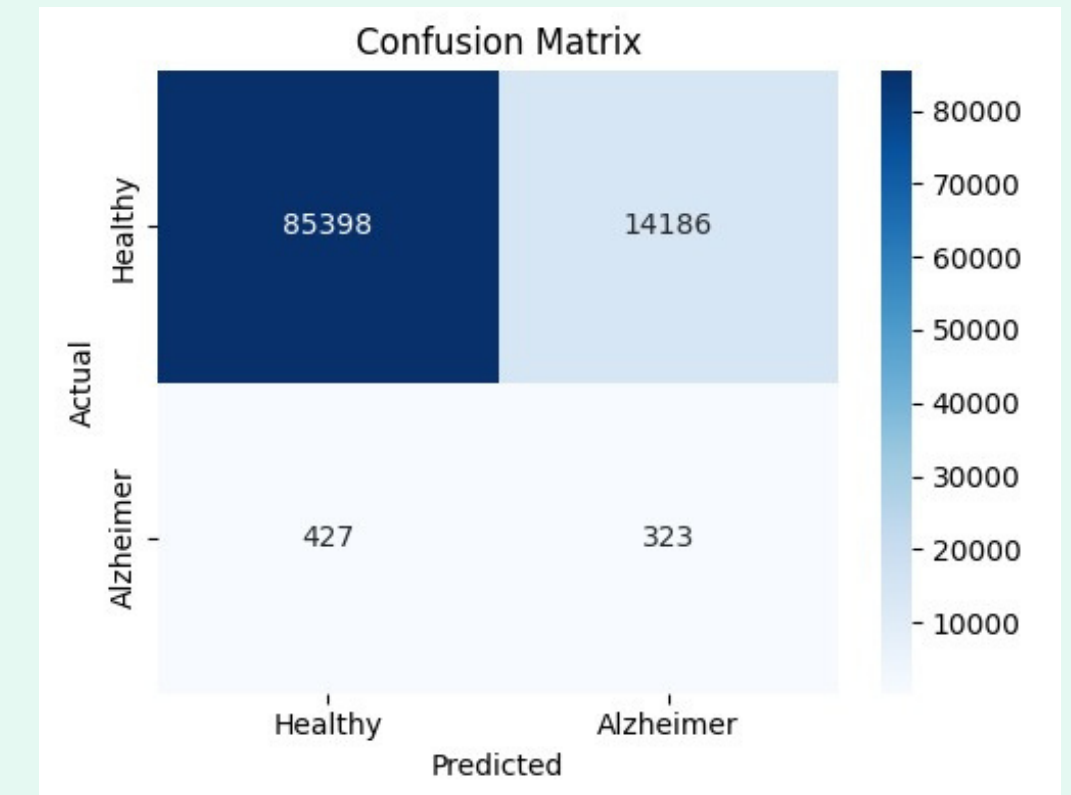
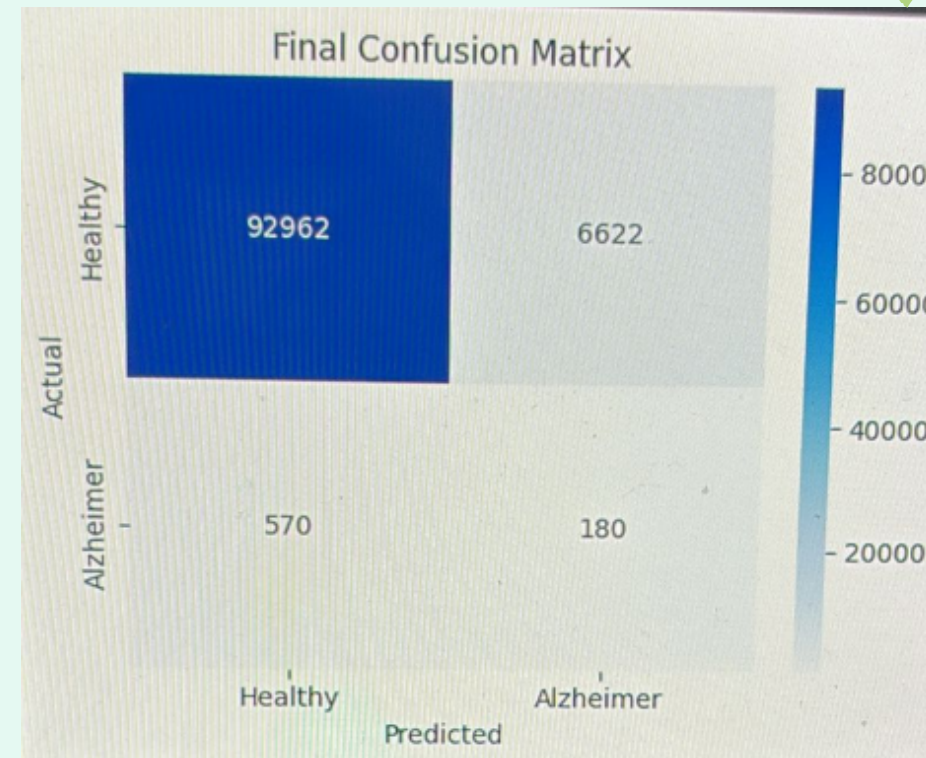
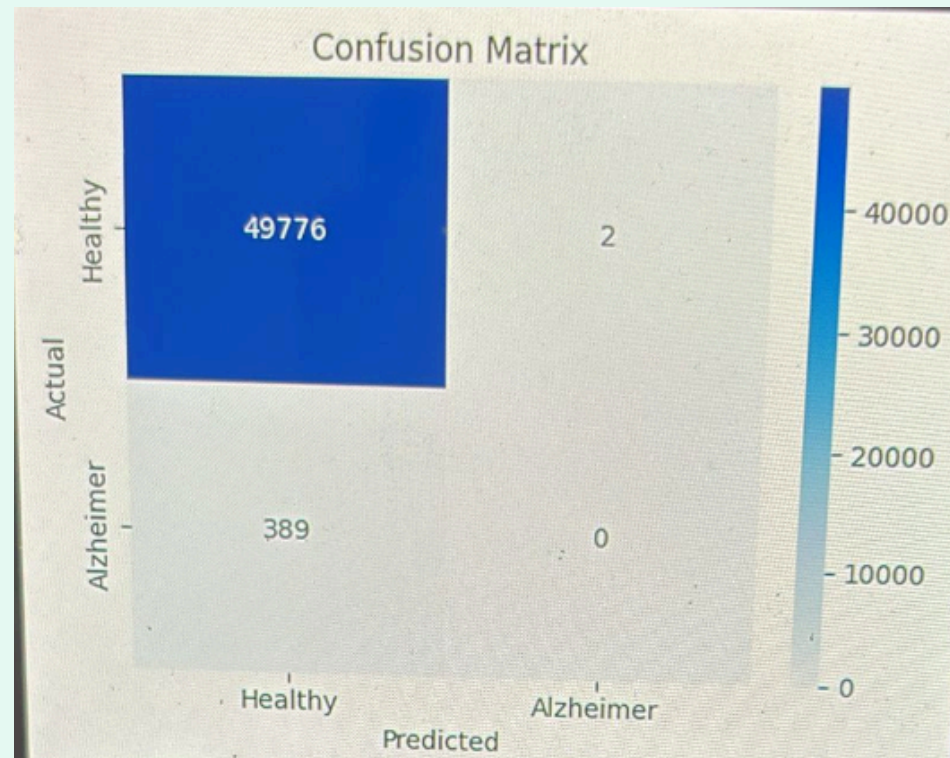
Each split is a decision.
More levels help the model understand
complex data,
but too many can cause overfitting.



Randomness in the model (subsample)



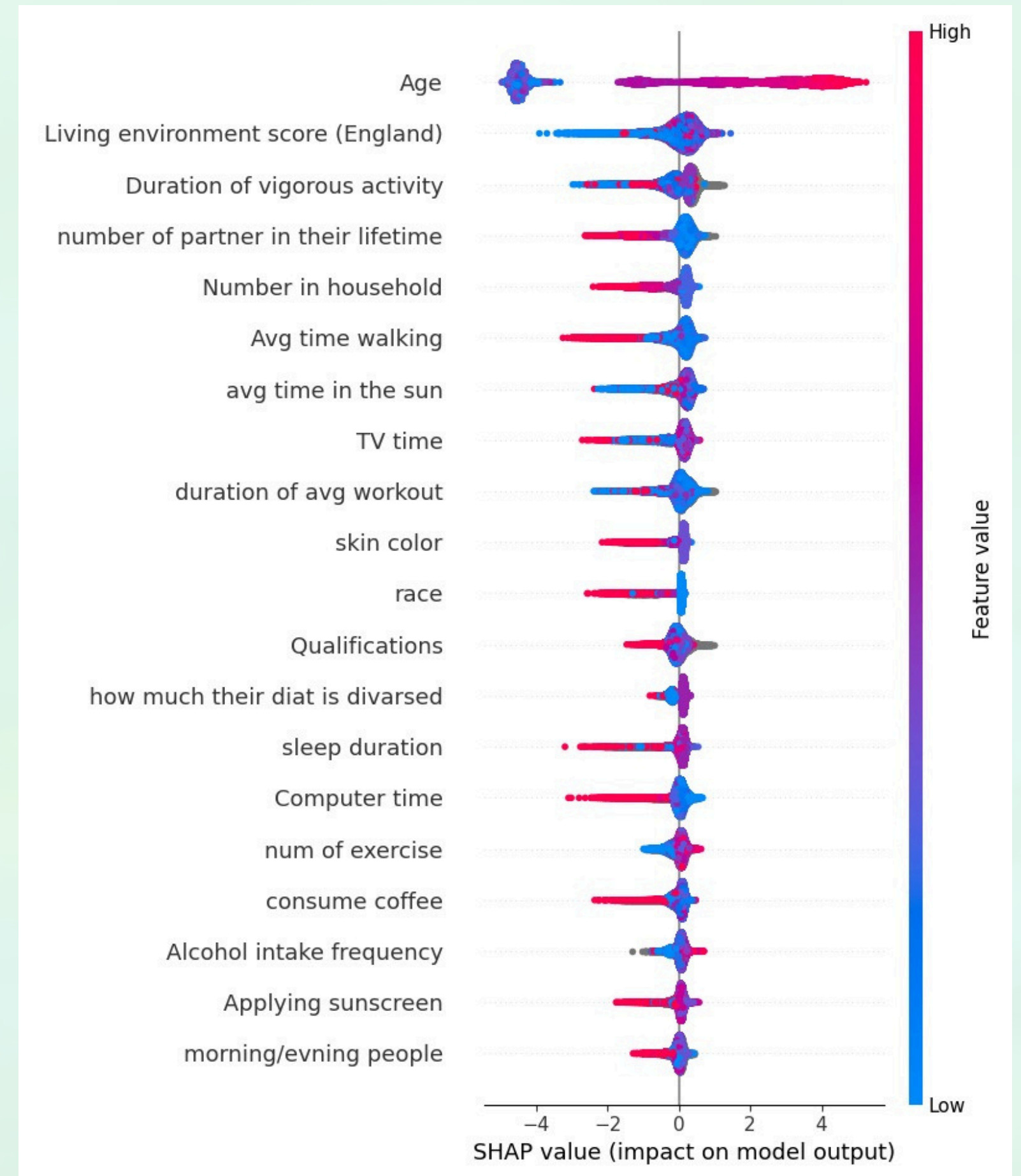
Improvment of the model



Model Evaluation: Feature Impact Analysis

SHAP Summary Plot

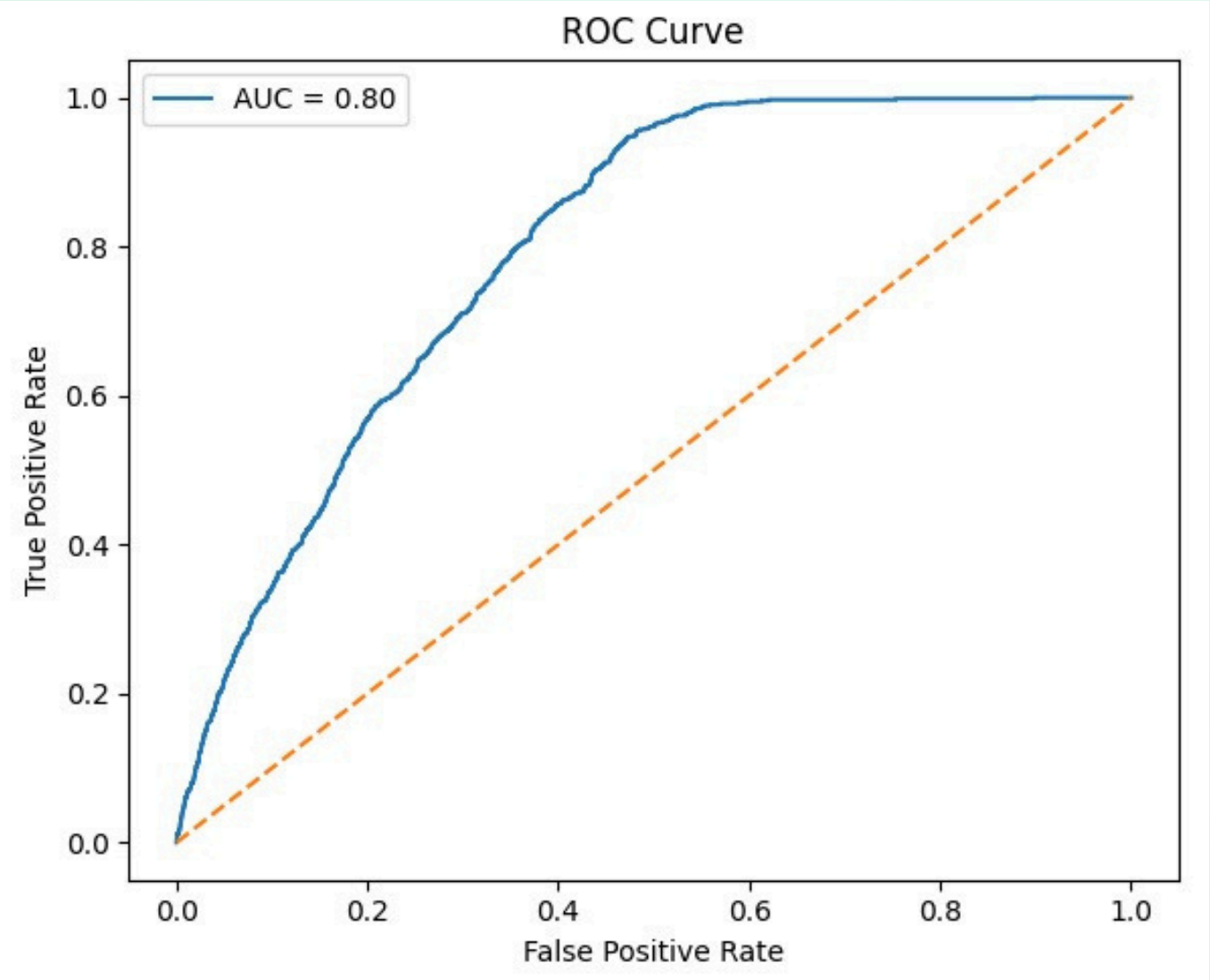
The plot shows how each feature affects predictions. Each dot is a prediction—red for high values, blue for low. Higher age values increase Alzheimer's risk, confirming age as a strong predictor



Model Performance Metrics

ROC Curve

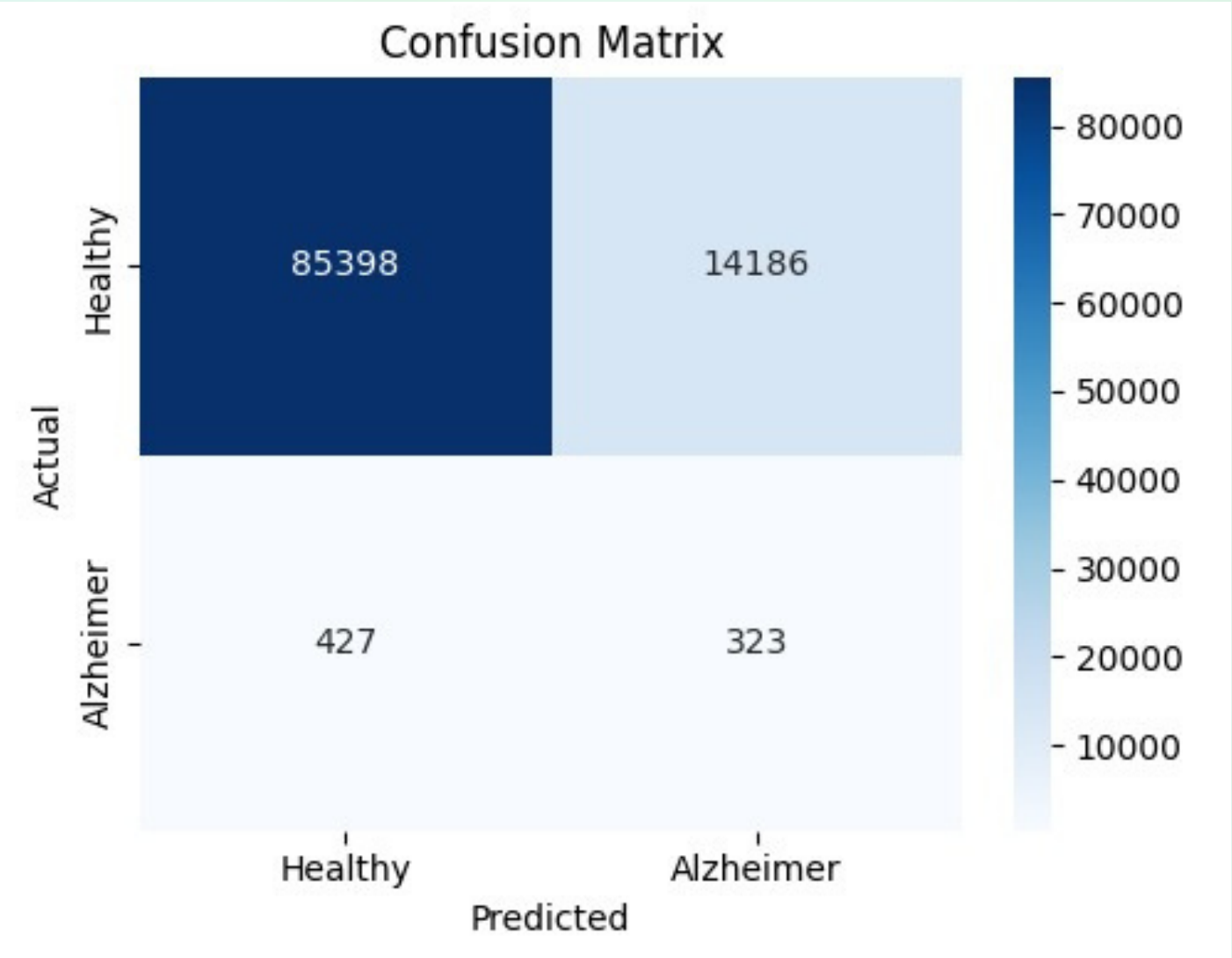
The ROC curve shows how well the model identifies Alzheimer’s patients while avoiding misclassifying healthy individuals. with an AUC of 0.80 indicating strong performance.



*AUC - area under the curve

Confusion Matrix

The confusion matrix shows the model accurately detects most healthy individuals but misses many Alzheimer’s cases, resulting in false negatives. This is common in imbalanced datasets where Alzheimer’s cases are rare.



Interview with Dr. Inbal Halevi Hochwald

Main disciplines: Medicine, Public Health, Nursing, and Geriatrics



Q: What is the treatment currently provided for patients with dementia?

Dr. Inbal Halevy Hochwald:

There is no cure for dementia. Current treatments, even the most advanced, can only slightly delay early symptoms and progression. Treatment today is supportive rather than curative.

Q: How do you see the collaboration between bioinformatics/statistics and clinical medicine?

Dr. Inbal Halevy Hochwald:

Medicine is ultimately quantitative, relying on statistics, probabilities, and generalizations. Quantitative research gives the essential data, while qualitative insights help explain the “why” behind the numbers. The best approach is a mixed-methods model that combines both.

Scan for the full interview



Q: Which behavioral factors influence dementia?

Dr. Inbal Halevy Hochwald:

Smoking, diabetes, and lack of physical activity worsen dementia, while loneliness is a major risk factor. Protective factors include social connections and a sense of meaning. Events like COVID-19 and war highlight how stress, displacement, and isolation can accelerate decline and even lead to premature death.

Q: Are social connections a major risk factor for dementia?

Dr. Inbal Halevy Hochwald:

Yes. Loneliness is a significant risk factor, while strong family ties and feeling loved and needed serve as important protective factors.

Q: Do you have recommendations that could help prevent or delay Alzheimer's?

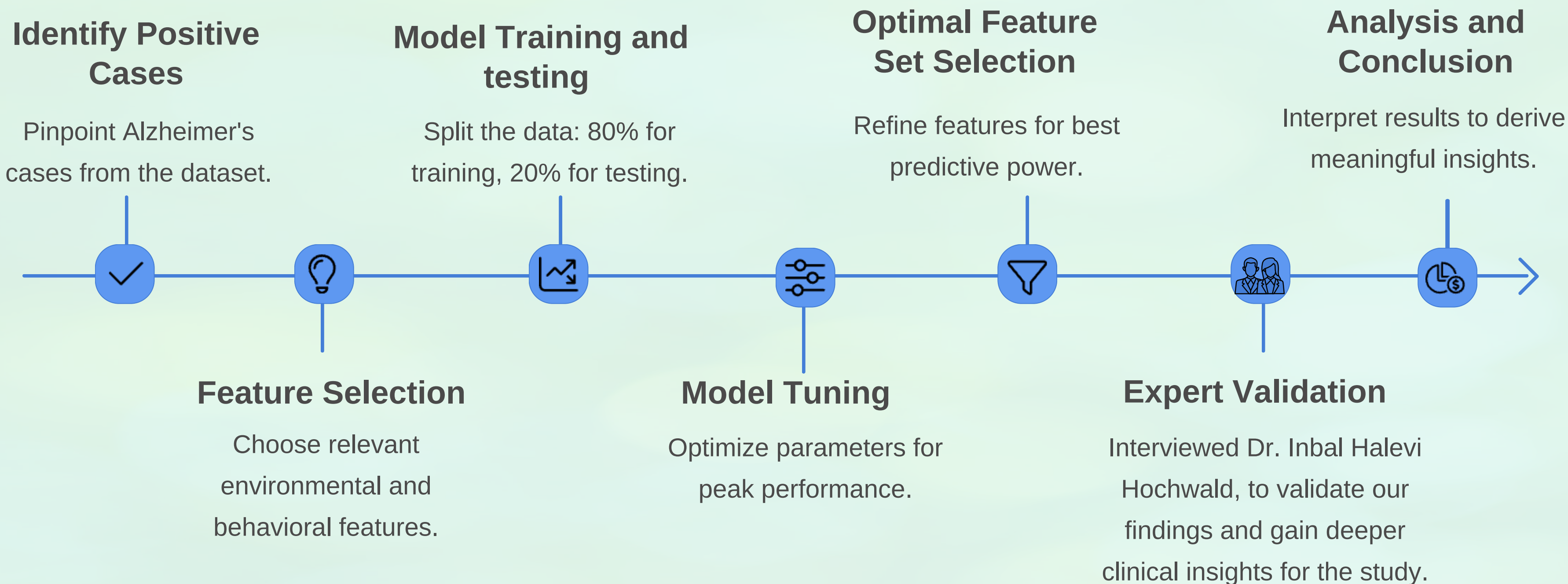
Dr. Inbal Halevy Hochwald:

There are three types of prevention: primary – before the disease, such as vaccines, secondary – early detection and lifestyle interventions; and tertiary – managing symptoms and preventing complications. Unfortunately, most resources go to tertiary care, while primary prevention is greatly underfunded.

Scan for the full
interview



Project Overview: A Step-by-Step Approach



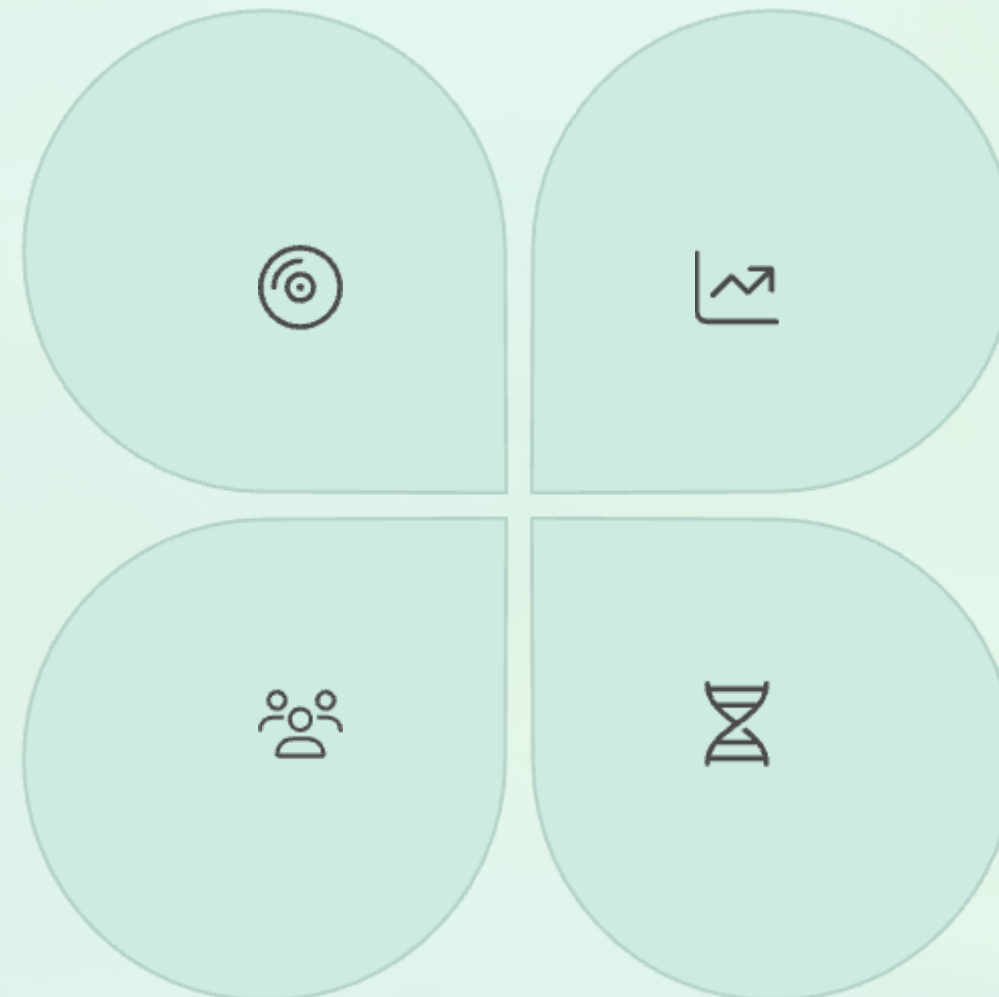
Conclusions and Future Directions

Research Impact

Advanced understanding of environmental risk factors associated with Alzheimer's disease

Population Benefit

Early identification of at-risk individuals improves intervention timing



Clinical Application

XGBoost model predicts risk with AUC=0.80, enabling targeted prevention strategies

Future Work

Will incorporate imaging and genetic data to enhance predictive power

Articles



- Digital dementia in the internet generation: excessive screen time during brain development will increase the risk of Alzheimer's disease and related dementias in adulthood
- <https://pubmed.ncbi.nlm.nih.gov/35164464/>



- Physical exercise in the prevention and treatment of Alzheimer's disease
- <https://pubmed.ncbi.nlm.nih.gov/32780691/>



- Sleep and Alzheimer's disease
- <https://pubmed.ncbi.nlm.nih.gov/24846773/>



- Smoking and increased Alzheimer's disease risk: a review of potential mechanisms
- <https://pubmed.ncbi.nlm.nih.gov/24924665/>



- Sex and the development of Alzheimer's disease
- <https://pubmed.ncbi.nlm.nih.gov/27870425/>



- Diet and Inflammation in Cognitive Ageing and Alzheimer's Disease
- <https://pubmed.ncbi.nlm.nih.gov/30949921/>