Table S1: Confusion matrix summarising the performance of a best classification model on PD data. The RF model trained using all 20183 genes in the dataset gave the best evaluation scores (accuracy = 0.702, ROC AUC = 0.743, prAUC = 0.762)

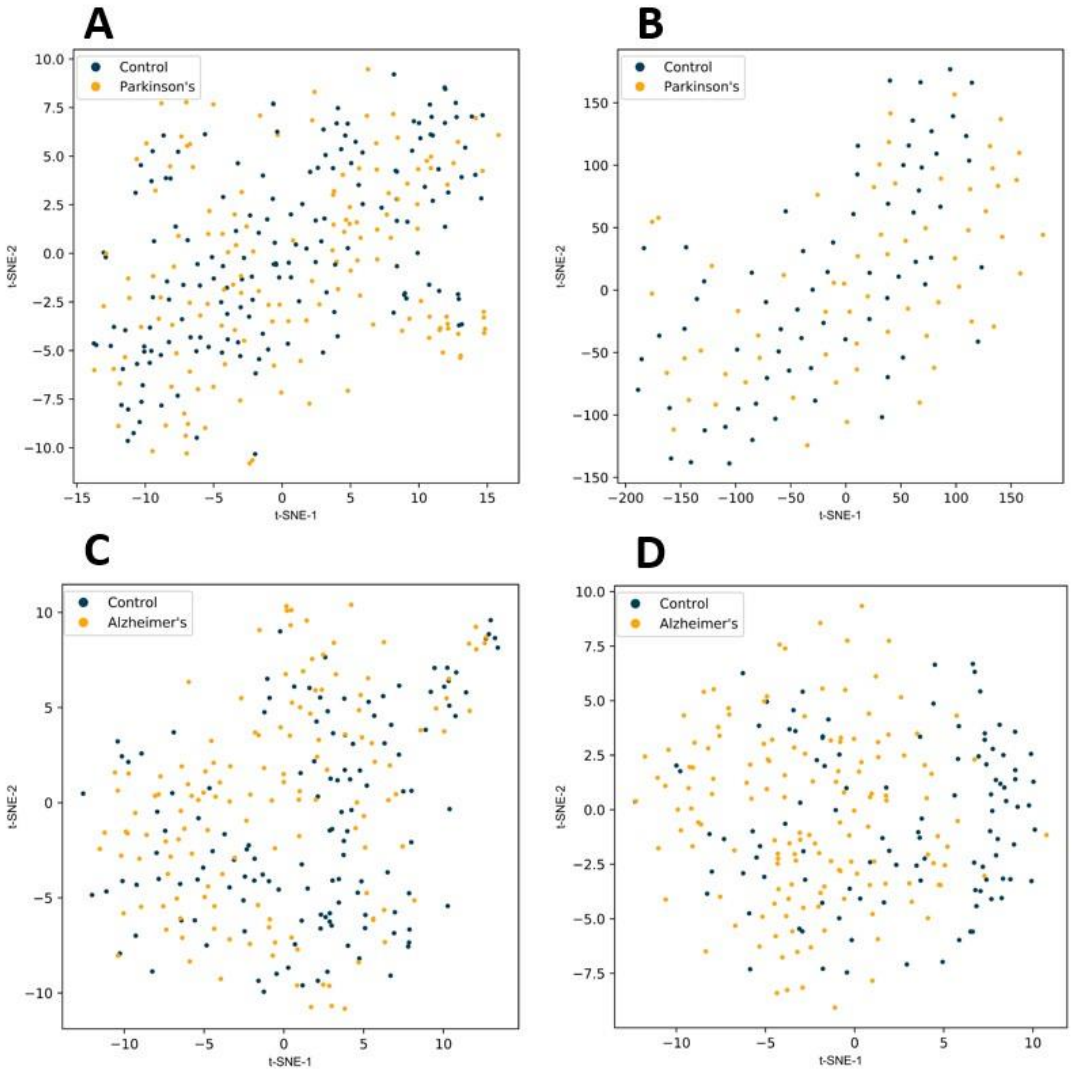
	True Positive	**True Negative**	**Total**
Predicted Positive	56	12	68
Predicted Negative	27	36	63
Total	83	48	131

Table S2: Confusion matrix summarising the performance of a best classification model on AD data. The RF model trained using the 159 features identified using VSSRFE gave the best evaluation scores (accuracy = 0.810, ROC AUC = 0.889, prAUC = 0.919)

	True Positive	**True Negative**	**Total**
Predicted Positive	83	21	104
Predicted Negative	26	117	143
Total	109	138	247

Table S3: Classification models and the parameters that were tuned on training data

Classification algorithm	Python library	Base python code	parameters tuned
LR	sklearn (28)	LogisticRegression(random state=2, class weight='balanced', penalty='l2', solver='liblinear')	• C
SVM with radial kernel	sklearn (28)	SVC(random state=142, kernel = 'rbf', class weight = 'balanced')	• C • Gamma
XGBoost	xgboost (17)	XGBClassifier(random state=42)	• scale_pos_weight • learning_rate • n_estimators • max_depth • min_child_weight • gamma • colsample_bytree • subsample • reg_alpha • reg_lambda
RF	sklearn (28)	RandomForestClassifier(random state=10, class weight='balanced')	• max_depth • min_samples_leaf • n_estimators • min_samples_split • max_features
MLP	sklearn (28)	MLPClassifier(random state=10, max iter = 10000, tol = 0.00001)	• activation • hidden_layer_size • solver
VAE	Keras (39)	model.compile(optimizer='adam', loss='categorical crossentropy', metrics=[metrics.AUC(name='PR', curve ='PR')])	• batch normalisation • dropout layers
CNN	Keras (39)	model.compile(loss = 'categorical crossentropy', optimizer = 'sgd', metrics = ['categorical accuracy'])	• dense layer size • filters • kernel size



Supplementary Figure S1: t-SNE plots for training and test data of PD and AD. PD training (A) and test (B) datasets and AD training (C) and test (D) datasets show no outliers and no clear distinction between disease and control samples.