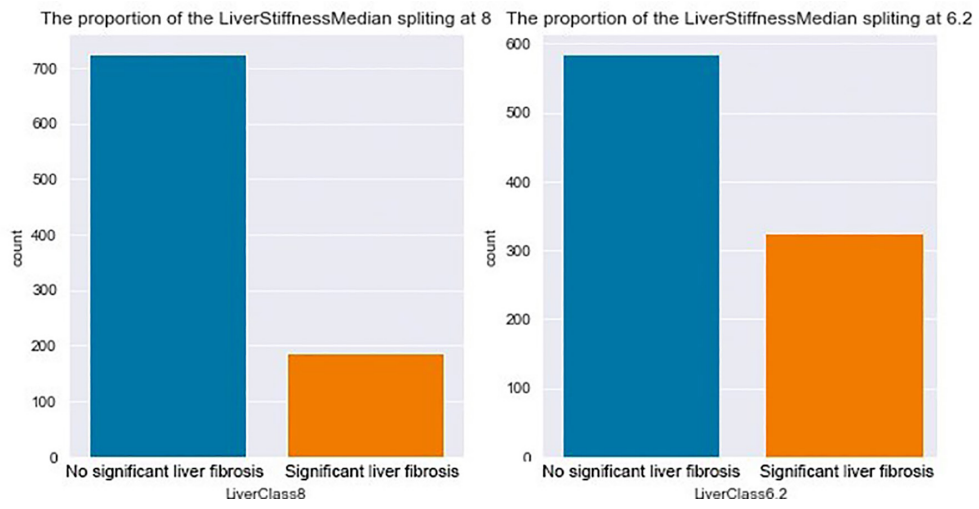
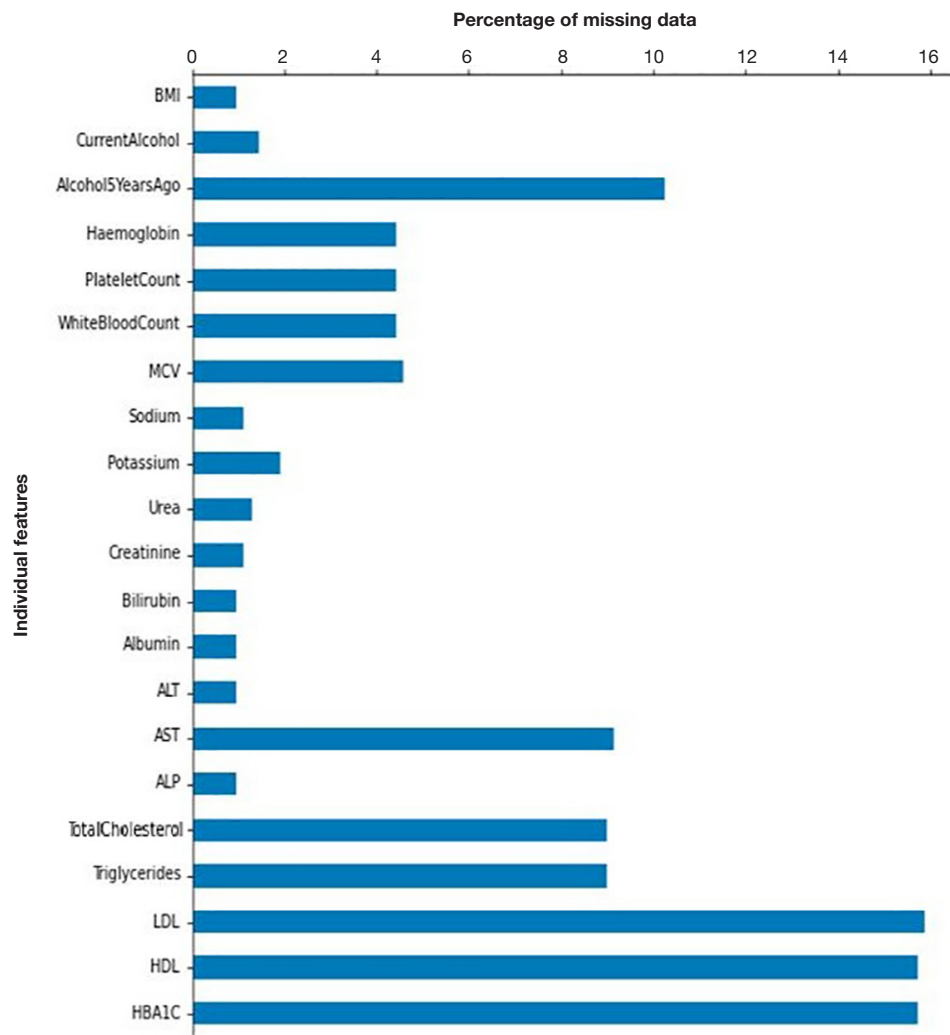


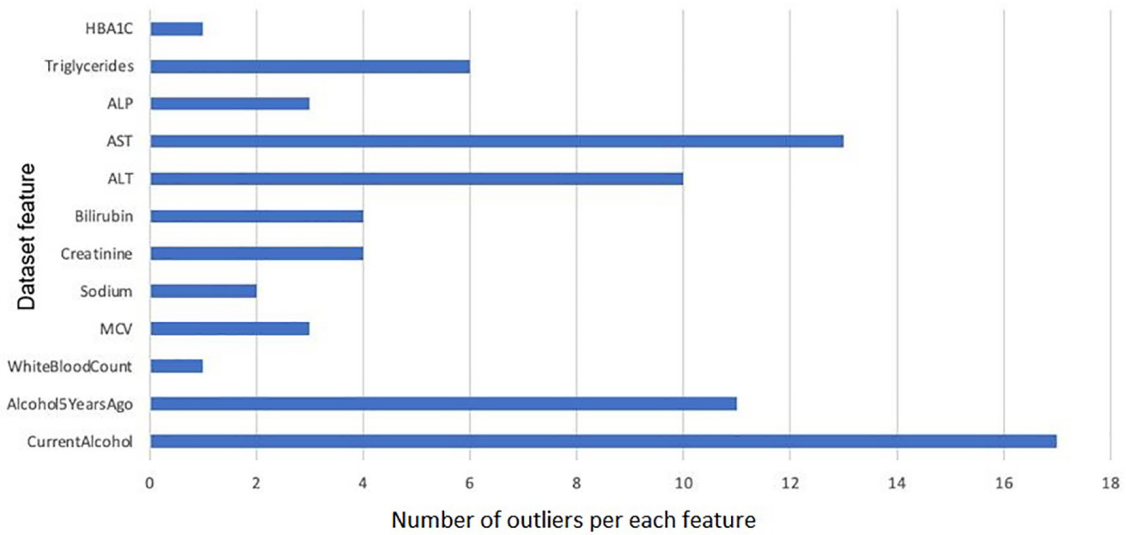
**Figure S1** Pearson correlation coefficient heat map for dataset features. BMI, body mass index; HbA1c, haemoglobin A1c; HDL, high density lipoprotein; ALT, alanine transaminase; AST, aspartate aminotransferase; MCV, mean corpuscular volume; ALP, alkaline phosphatase.



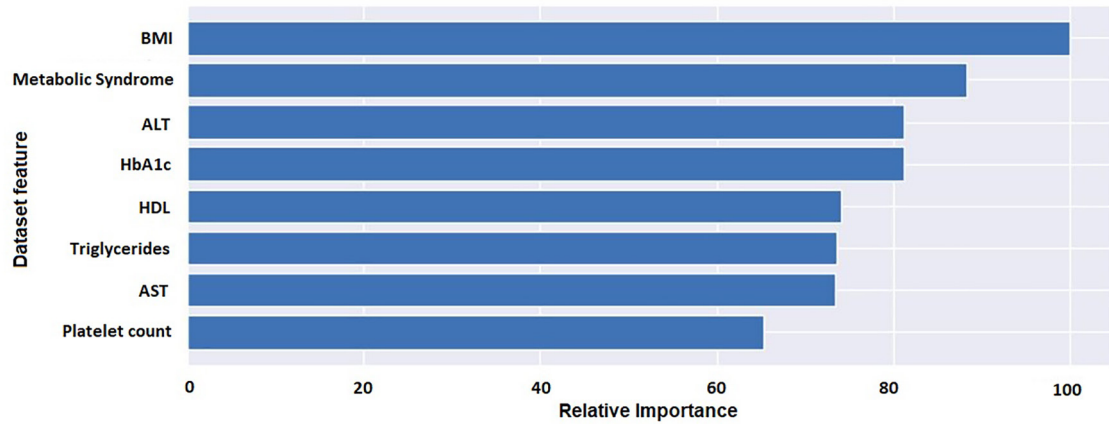
**Figure S2** Classification imbalance.



**Figure S3** Missing value percentage for individual features. BMI, body mass index; HbA1c, haemoglobin A1c; HDL, high density lipoprotein; ALT, alanine transaminase; AST, aspartate aminotransferase; MCV, mean corpuscular volume; ALP, alkaline phosphatase; LDL, low density lipoprotein.



**Figure S4** Outliers detected in the dataset with Q1 0.5 and Q3 0.95. HbA1c, haemoglobin A1c; ALT, alanine transaminase; AST, aspartate aminotransferase; MCV, mean corpuscular volume; ALP, alkaline phosphatase.



**Figure S5** Importance of the features after applying Random Forest Tree. BMI, body mass index; HbA1c, haemoglobin A1c; HDL, high density lipoprotein; ALT, alanine transaminase; AST, aspartate aminotransferase.