We claim:

1. A method comprising:

a) acquiring a class-imbalanced data set, wherein the class-imbalanced data set comprises biological data from a plurality of subjects, wherein the biological data of each subject includes an observation, a time value, and a plurality of clinical measurements, and wherein the biological data is categorized as being part of a majority data class or a minority data class, wherein the majority data class has a greater number of observations than the minority data class;

b) downsampling the class-imbalanced data set to generate a downsampled data set, wherein the downsampling results in the majority data class having an equivalent or substantially equivalent number of observations as the minority data class; and

c) performing cross-validation on the downsampled data set with a survival analysis to generate a survival model;

wherein the observation comprises an event or no event at a specific time value; and

wherein an AUC, sensitivity, specificity, and/or C-index of the survival model is closer to 1 than a AUC, sensitivity, specificity, and/or C-index of a survival model where the class-imbalanced data set was not downsampled prior to the survival analysis.

2. The method of claim 1, wherein the class-imbalanced data set is a survival data set.

3. The method of claim 1, wherein the event is a disease, disorder, or condition of a subject.

4. The method of claim 1, wherein the survival analysis is selected from the group consisting of a Cox proportional hazard analysis, a random forest analysis, an accelerated failure time analysis, and any combination thereof.

5. The method of claim 4, further comprising an elastic net penalty.

6. The method of claim 1, wherein the cross-validation is at least a 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 11-fold, 12-fold, 13-fold, 14-fold, 15-fold, 16-fold, 17-fold, 18-fold, 19-fold, or 20-fold cross-validation.

7. The method of claim 1, wherein the survival model comprises from 5 to 1,000 features, wherein each feature is selected from the group consisting of a protein measurement, a clinical factor, and a combination thereof.

8. The method of claim 7, wherein the clinical factor is selected from the group consisting of age, weight, blood pressure, height, BMI, cholesterol, sex, and a combination thereof.

9. The method of claim 1, wherein the clinical measurements are selected from proteomic measurements, genomic measurements, transcriptome measurements, metabolomics measurements, or a combination thereof.

10. The method of claim 1, wherein the cross-validation is selected from k-fold cross-validation, Monte Carlo cross-validation, and Leave N Out validation.

11. The method of claim 1, wherein the majority data class is 95% of the class-imbalanced data set and the minority data class is 5% of the class-imbalance data set.

12. The method of claim 1, wherein the majority data class is 90% of the class-imbalanced data set and the minority data class is 10% of the class-imbalance data set.

13. The method of claim 1, wherein the majority data class is 85% of the class-imbalanced data set and the minority data class is 15% of the class-imbalance data set.

14. The method of claim 1, wherein the majority data class is 80% of the class-imbalanced data set and the minority data class is 20% of the class-imbalance data set.

15. The method of claim 1, wherein the majority data class is 75% of the class-imbalanced data set and the minority data class is 25% of the class-imbalance data set.

16. The method of claim 1, wherein the majority data class is 70% of the class-imbalanced data set and the minority data class is 30% of the class-imbalance data set.

17. The method of claim 1, wherein the majority data class is 65% of the class-imbalanced data set and the minority data class is 35% of the class-imbalance data set.

18. The method of claim 1, wherein the majority data class is 60% of the class-imbalanced data set and the minority data class is 40% of the class-imbalance data set.

19. A method comprising:

a) downsampling a class-imbalanced data set to generate a downsampled data set, wherein the downsampling results in a majority data class having an equivalent or substantially equivalent number of observations as a minority data class; and

b) performing cross-validation on the downsampled data set with a survival analysis to generate a survival model;

wherein the observation comprises an event or no event at a specific time value;

wherein the class-imbalanced data set comprises biological data from a plurality of subjects, wherein the biological data of each subject includes an observation, a time value, and a plurality of clinical measurements, and wherein the biological data is categorized as being part of the majority data class or the minority data class, wherein the majority data class has a greater number of observations than the minority data class; and

wherein an AUC, sensitivity, specificity, and/or C-index of the survival model is closer to 1 than a AUC, sensitivity, specificity, and/or C-index of a survival model where the class-imbalanced data set was not downsampled prior to the survival analysis.

20. The method of claim 19, wherein the AUC is calculated based on the determination of whether or not a subject will have an event by a specified time-point.

21. A computer-implemented method for determining disease risk comprising:

a) acquiring a class-imbalanced data set, wherein the class-imbalanced data set comprises biological data from a plurality of subjects, wherein the biological data of each subject includes an observation, a time value and a plurality of clinical measurements, and wherein the biological data is categorized as being part of a majority data class or a minority data class, wherein the majority data class has a greater number of observations than the minority data class;

b) downsampling the class-imbalanced data set to generate a downsampled data set, wherein the downsampling results in the majority data class having an equivalent or substantially equivalent number of observations as the minority data class; and

c) performing cross-validation on the downsampled data set with a survival analysis to generate a survival model;

wherein the observation comprises an event or no event at a specific time value;

wherein step b) and step c) are computed with a computer system; and

wherein an AUC, sensitivity, specificity, and/or C-index of the survival model is closer to 1 than a AUC, sensitivity, specificity, and/or C-index of a survival model where the class-imbalanced data set was not downsampled prior to the survival analysis.

22. A program storage device readable by a computer, tangibly embodying a program of instructions executable by the computer to perform the method steps for a method for determining disease risk comprising:

a) acquiring a class-imbalanced data set, wherein the class-imbalanced data set comprises biological data from a plurality of subjects, wherein the biological data of each subject includes an observation, a time value and a plurality of clinical measurements, and wherein the biological data is categorized as being part of a majority data class or a minority data class, wherein the majority data class has a greater number of observations than the minority data class;

b) downsampling the class-imbalanced data set to generate a downsampled data set, wherein the downsampling results in the majority data class having an equivalent or substantially equivalent number of observations as the minority data class; and

c) performing cross-validation on the downsampled data set with a survival analysis to generate a survival model;

wherein the observation comprises an event or no event at a specific time value; and

wherein an AUC, sensitivity, specificity, and/or C-index of the survival model is closer to 1 than a AUC, sensitivity, specificity, and/or C-index of a survival model where the class-imbalanced data set was not downsampled prior to the survival analysis.

23. A computing system for determining disease risk comprising: a memory for storing programmed instructions; a processor configured to execute the programmed instructions to perform operations comprising:

a) acquiring a class-imbalanced data set, wherein the class-imbalanced data set comprises biological data from a plurality of subjects, wherein the biological data of each subject includes an observation, a time value and a plurality of clinical measurements, and wherein the biological data is categorized as being part of a majority data class or a minority data class, wherein the majority data class has a greater number of observations than the minority data class;

b) downsampling the class-imbalanced data set to generate a downsampled data set, wherein the downsampling results in the majority data class having an equivalent or substantially equivalent number of observations as the minority data class; and

c) performing cross-validation on the downsampled data set with a survival analysis to generate a survival model;

wherein the observation comprises an event or no event at a specific time value; and

wherein an AUC, sensitivity, specificity, and/or C-index of the survival model is closer to 1 than a AUC, sensitivity, specificity, and/or C-index of a survival model where the class-imbalanced data set was not downsampled prior to the survival analysis.

24. A non-transitory, computer readable media with instructions stored thereon that are executable by a processor to perform operations of:

a) acquiring a class-imbalanced data set, wherein the class-imbalanced data set comprises biological data from a plurality of subjects, wherein the biological data of each subject includes an observation, a time value, and a plurality of clinical measurements, and wherein the biological data is categorized as being part of a majority data class or a minority data class,

wherein the majority data class has a greater number of observations than the minority data class;

b) downsampling the class-imbalanced data set to generate a downsampled data set, wherein the downsampling results in the majority data class having an equivalent or substantially equivalent number of observations as the minority data class; and

c) performing cross-validation on the downsampled data set with a survival analysis to generate a survival model;

wherein the observation comprises an event or no event at a specific time value; and

wherein an AUC, sensitivity, specificity, and/or C-index of the survival model is closer to 1 than a AUC, sensitivity, specificity, and/or C-index of a survival model where the class-imbalanced data set was not downsampled prior to the survival analysis.

25. A computer-implemented method for determining disease risk comprising:

a) receiving with a computer a class-imbalanced data set, wherein the class-imbalanced data set comprises biological data from a plurality of subjects, wherein the biological data of each subject includes an observation, a time value and a plurality of clinical measurements, and wherein the biological data is categorized as being part of a majority data class or a minority data class, wherein the majority data class has a greater number of observations than the minority data class;

b) downsampling with the computer the class-imbalanced data set to generate a downsampled data set, wherein the downsampling results in the majority data class having an equivalent or substantially equivalent number of observations as the minority data class; and

c) performing with the computer cross-validation on the downsampled data set with a survival analysis to generate a survival model;

wherein the observation comprises an event or no event at a specific time value; and

wherein an AUC, sensitivity, specificity, and/or C-index of the survival model is closer to 1 than a AUC, sensitivity, specificity, and/or C-index of a survival model where the class-imbalanced data set was not downsampled prior to the survival analysis.

26. The method of claim 19, 21 or 25, the device of claim 22, the system of claim 23 or the computer readable media of claim 24, wherein the class-imbalanced data set is a survival data set.

27. The method of claim 19, 21 or 25, the device of claim 22, the system of claim 23 or the computer readable media of claim 24, wherein the event is a disease, disorder, or condition of a subject.

28. The method of claim 19, 21 or 25, the device of claim 22, the system of claim 23 or the computer readable media of claim 24, wherein the survival analysis is selected from the group consisting of a Cox proportional hazard analysis, a random forest analysis, an accelerated failure time analysis, and any combination thereof.

29. The method, device, system or computer readable media of claim 28, further comprising an elastic net penalty.

30. The method of claim 19, 21 or 25, the device of claim 22, the system of claim 23 or the computer readable media of claim 24, wherein the cross-validation is at least a 2-fold, 3-fold, 4-fold, 5-fold, 6-fold, 7-fold, 8-fold, 9-fold, 10-fold, 11-fold, 12-fold, 13-fold, 14-fold, 15-fold, 16-fold, 17-fold, 18-fold, 19-fold, or 20-fold cross-validation.

31. The method of claim 19, 21 or 25, the device of claim 22, the system of claim 23 or the computer readable media of claim 24, wherein the survival model comprises from 5 to 1,000

features, wherein each feature is selected from the group consisting of a protein measurement, a clinical factor, and a combination thereof.

32. The method, device, system or computer readable media of claim 31, wherein the clinical factor is selected from the group consisting of age, weight, blood pressure, height, BMI, cholesterol, sex, and a combination thereof.

33. The method of claim 19, 21 or 25, the device of claim 22, the system of claim 23 or the computer readable media of claim 24, wherein the clinical measurements are selected from proteomic measurements, genomic measurements, transcriptome measurements, metabolomics measurements, or a combination thereof.

34. The method of claim 19, 21 or 25, the device of claim 22, the system of claim 23 or the computer readable media of claim 24, wherein the cross-validation is selected from k-fold cross-validation, Monte Carlo cross-validation, and Leave N Out validation.

35. The method of claim 19, 21 or 25, the device of claim 22, the system of claim 23 or the computer readable media of claim 24, wherein the majority data class is 95% of the class-imbalanced data set and the minority data class is 5% of the class-imbalance data set.

36. The method of claim 19, 21 or 25, the device of claim 22, the system of claim 23 or the computer readable media of claim 24, wherein the majority data class is 90% of the class-imbalanced data set and the minority data class is 10% of the class-imbalance data set.

37. The method of claim 19, 21 or 25, the device of claim 22, the system of claim 23 or the computer readable media of claim 24, wherein the majority data class is 85% of the class-imbalanced data set and the minority data class is 15% of the class-imbalance data set.

38. The method of claim 19, 21 or 25, the device of claim 22, the system of claim 23 or the computer readable media of claim 24, wherein the majority data class is 80% of the class-imbalanced data set and the minority data class is 20% of the class-imbalance data set.

39. The method of claim 19, 21 or 25, the device of claim 22, the system of claim 23 or the computer readable media of claim 24, wherein the majority data class is 75% of the class-imbalanced data set and the minority data class is 25% of the class-imbalance data set.

40. The method of claim 19, 21 or 25, the device of claim 22, the system of claim 23 or the computer readable media of claim 24, wherein the majority data class is 70% of the class-imbalanced data set and the minority data class is 30% of the class-imbalance data set.

41. The method of claim 19, 21 or 25, the device of claim 22, the system of claim 23 or the computer readable media of claim 24, wherein the majority data class is 65% of the class-imbalanced data set and the minority data class is 35% of the class-imbalance data set.

42. The method of claim 19, 21 or 25, the device of claim 22, the system of claim 23 or the computer readable media of claim 24, wherein the majority data class is 60% of the class-imbalanced data set and the minority data class is 40% of the class-imbalance data set.