*8/21/2023 Machine Learning Analysis Results with No Down Syndrome, No APML*

Big Conclusions:

* We can include hemoglobin and monocytes because there’s only a 3 count difference from our normal models.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Total Admissions** | **Neg Admissions** | **Pos Admissions** |
| **Original Dataset** | 624 | 522 | 102 (16.3%) |
| **No Down Syndrome, No APML** | 480 | 389 | 91 (18.9%) |
| **All Cytarabine** | 429 | 343 | 86 (20.0%) |
| **All Neutrophil, All Platelet** | 426 | 340 | 86 (20.2%) |
| **All Hemoglobin, All Monocytes** | 423 | 338 | 85 (20.1%) |

* The best performing machine learning model is Logistic Regression with a Cytarabine mg/m2/day threshold of 2000. The baseline Fever model still performs just as well. This model also performs slightly better than before when we had all 624 admissions (0.741 vs. 0.737).
* Including hemoglobin and monocytes increases the accuracy (0.72 🡪 0.74)
* Using a threshold of 2000 mg/m2/day is better than having three thresholds and better than having no threshold (just the cytarabine dosage itself)
* There seems to be a very slight improvement using 2000 as a threshold versus 1000.
* The most important features of the random forest model are:
  1. Max\_temp\_38.5
  2. Lowest\_neutrophil
  3. Port
  4. Cyt\_2000
  5. Lowest\_platelet

**Figure 1.**

A graph of a graph

Description automatically generated with medium confidence

Table 1: Patient Demographics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **Negative BSI** | **Positive BSI** | **p-values** | **Total** |
| Total number of patients | | 43 (45.3%) | 52 (54.7%) | 0.000 | 95 |
| Sex | Male | 20 (46.5%) | 27 (51.9%) | 0.604 | 47 (49.5%) |
|  | Female | 23 (53.5%) | 25 (48.1%) |  | 48 (50.5%) |
| Race/Ethnicity | White/Caucasian | 30 (69.8%) | 39 (75.0%) | 0.574 | 69 (72.6%) |
|  | Non-white & undefined | 13 (30.2%) | 13 (25.0%) |  | 26 (27.4%) |
| Diagnosis age | Median (range) | 6.76 (0.05-18.51) | 8.86 (0.34-18.75) | 0.672 | 7.98 (0.05-18.75) |
| (years) | Interquartile range | (1.75-12.38) | (2.04-12.88) |  | (1.87-13.02) |
| Age at 1st | Median (range) |  | 9.20 (0.93-19.22) |  |  |
| infection (years) | Interquartile range |  | (2.06-13.41) |  |  |
| Diagnosis | AML | 37 (86.0%) | 47 (90.4%) | 0.517 | 84 (88.4%) |
|  | 2nd AML | 6 (14.0%) | 5 (9.6%) | 0.517 | 11 (11.6%) |
| Number of | Median (range) | 4 (1-8) | 6 (1-13) |  | 5 (1-16) |
| Admissions | Average (95% CI) | 4.07 (3.42-4.72) | 5.87 (5.22-6.51) |  | 5.05 (4.56-5.54) |
| Deaths |  | 18 (41.9%) | 18 (34.6%) | 0.474 | 36 (37.9%) |

Table 2: Admissions Demographics

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | **BSI Negative** | **BSI Positive** | **p-values** | **Total** |
| **Total number of Admissions** | | 389 (80.9%) | 91 (18.9%) |  | 480 |
| **Sex** | Male | 199 (51.2%) | 54 (59.3%) | 0.160 | 253 (52.7%) |
|  | Female | 190 (48.8%) | 37 (40.7%) |  | 227 (47.3%) |
| **Race** | White/Caucasian | 265 (68.1%) | 69 (75.8%) | 0.151 | 334 (69.6%) |
|  | Non-white & undefined | 124 (31.9%) | 22 (24.2%) |  | 146 (30.4%) |
| **Diagnosis** | AML | 371 (95.4%) | 85 (93.4%) | 0.440 | 456 (95.0%) |
|  | 2nd AML | 18 (4.6%) | 6 (6.6%) | 0.440 | 24 (5.0%) |
| **Age at Admission (years)** | Median(range) | 7.15 (0-19.20) | 9.81 (0.93-19.22) | 0.021 | 8.31 (0-19.22) |
| **First BMI (kg/m^2)** | Median(range) | 17.40 (10-44.60) | 17.9 (13.50-46.60) | 0.065 | 17.50 (10-46.60) |
| **Number of Neutropenic Admissions** | | 319 (82.0%) | 88 (96.7%) | 0.000 | 370 (84.8%) |
| **Lowest ANC** | Median(range) | 0.008 (0-17.978) | 0.003 (0-13.899) | 0.058 | 0.007 (0-17.978) |
| **Lowest Platelet count** | Median(range) | 27 (0-1176) | 11 (1-154) | 0.002 | 15 (0-1176) |
| **LOS (days)** | Median(range) | 24.29 (0.35-74.30) | 28.40 (2.19-81.86) | 0.000 | 25.23 (0.35-81.86) |
| **Number of PICU Visits** | | 34 (8.7%) | 20 (22.0%) | 0.000 | 54 (11.3%) |
| **PICU LOS (days)** | Median(range) | 2.94 (0.04-41.03) | 3.16 (0.00-48.85) | 0.456 | 3.08 (0.00-48.85) |
| **Total number of admissions with medication data\*** | | **343 (80.0%)** | **86 (20.0%)** |  | **429 (89.4%)** |
| **Cytarabine** | With | 237 (60.9%) | 65 (71.4%) | 0.24 | 302 (70.4%) |
|  | Without | 106 (27.2%) | 21 (23.1%) |  | 127 (29.6%) |
| **Levofloxacin** | With | 142 (36.5%) | 21 (23.1%) | 0.004 | 163 (38.0%) |
|  | Without | 247 (63.5%) | 70 (76.9%) |  | 266 (62.0%) |
| **Vancomycin** | With | 293 (75.3%) | 74 (81.3%) | 0.682 | 363 (84.6%) |
|  | Without | 96 (24.7%) | 17 (18.7%) |  | 66 (15.4%) |

Table 3: Logistic Regression Analysis

|  |  |  |  |
| --- | --- | --- | --- |
|  | (1) | (2) | (3) |
|  | Cytarabine | Levofloxacin | All |
| **Age** | 1.037 | 1.037 | 1.030 |
| (0.992, 1.083) | (0.991, 1.084) | (0.984, 1.078) |
| **Male** | 1.392 | 1.221 | 1.253 |
| (0.833, 2.326) | (0.725, 2.054) | (0.739, 2.125) |
| **White/Caucasian** | 1.422 | 1.558 | 1.534 |
| (0.821, 2.462) | (0.896, 2.707) | (0.877, 2.680) |
| **Cytarabine** | 1.368 |  | 2.039\*\* |
| (0.787, 2.375) |  | (1.129, 3.684) |
| **Levo** |  | 0.472\*\*\* | 0.354\*\*\* |
|  | (0.273, 0.815) | (0.197, 0.636) |
| **Psuedo R-squared** | 0.024 | 0.039 | 0.053 |

\*p < 0.10, \*\*p < 0.05, \*\*\*p < 0.01.

Table 4: Performance of Random Forest and Logistic Regression model against baseline models ‘Fever’ and ‘Neutropenia’ at the same Sensitivity.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Proportion Positive** | **Sensitivity** | **Specificity** | **Positive Predictive Value** | **Negative Predictive Value** |
| **Fever** | 0.68 | 0.93 | 0.38 | 0.27 | 0.96 |
| **Logistic Regression** | 0.73 | 0.93 | 0.32 | 0.25 | 0.95 |
| **Random Forest** | 0.74 | 0.93 | 0.31 | 0.25 | 0.95 |
| **Neutropenia** | 0.87 | 0.96 | 0.16 | 0.22 | 0.95 |