

Abstract:

```

1  # ===== 1. Import Libraries =====
2  import pandas as pd
3  import numpy as np
4  import matplotlib.pyplot as plt
5  import seaborn as sns
6  from sklearn.preprocessing import StandardScaler, MinMaxScaler, RobustScaler
7  from sklearn.model_selection import train_test_split, cross_val_score, GridSearchCV
8  from sklearn.metrics import accuracy_score, confusion_matrix, roc_auc_score, f1_score, precision_score, recall_score
9  from sklearn.linear_model import LogisticRegression
10 from sklearn.svm import SVC
11 from sklearn.tree import DecisionTreeClassifier
12 from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
13 from sklearn.neighbors import KNeighborsClassifier
14 from sklearn.naive_bayes import GaussianNB
15 from sklearn.metrics import classification_report
16
17 # ===== 2. Load Data =====
18 # Assuming data is loaded from a CSV file named 'data.csv'
19 data = pd.read_csv('data.csv')
20
21 # ===== 3. Data Preprocessing =====
22 # Check for missing values
23 data.isnull().sum()
24
25 # Drop missing values
26 data = data.dropna()
27
28 # Encode categorical variables
29 data = pd.get_dummies(data, drop_first=True)
30
31 # Split data into features (X) and target variable (y)
32 X = data.drop('target', axis=1)
33 y = data['target']
34
35 # Scale features
36 scaler = StandardScaler()
37 X_scaled = scaler.fit_transform(X)
38
39 # Split data into training and testing sets
40 X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42)
41
42 # ===== 4. Model Training =====
43 # Define models
44 models = {
45     'Logistic Regression': LogisticRegression(),
46     'Support Vector Machine': SVC(),
47     'Decision Tree': DecisionTreeClassifier(),
48     'Random Forest': RandomForestClassifier(),
49     'Gradient Boosting': GradientBoostingClassifier(),
50     'K-Nearest Neighbors': KNeighborsClassifier(),
51     'Naive Bayes': GaussianNB()
52 }
53
54 # Train models
55 for model_name, model in models.items():
56     model.fit(X_train, y_train)
57
58 # ===== 5. Model Evaluation =====
59 # Evaluate models on the test set
60 for model_name, model in models.items():
61     y_pred = model.predict(X_test)
62     accuracy = accuracy_score(y_test, y_pred)
63     f1 = f1_score(y_test, y_pred)
64     precision = precision_score(y_test, y_pred)
65     recall = recall_score(y_test, y_pred)
66     print(f'{model_name} Accuracy: {accuracy}, F1 Score: {f1}, Precision: {precision}, Recall: {recall}')
67
68 # ===== 6. Feature Importance =====
69 # Feature importance for Random Forest
70 rf_model = models['Random Forest']
71 feature_importance = rf_model.feature_importances_
72
73 # Plot feature importance
74 plt.bar(X.columns, feature_importance)
75 plt.title('Feature Importance for Random Forest')
76 plt.show()
77
78 # ===== 7. Model Interpretation =====
79 # Using SHAP for model interpretation
80 from shap import explain
81
82 # Explain a single prediction
83 shap_explainer = explain(rf_model)
84 shap_values = shap_explainer.shap_values(X_test)
85 shap_summary_plot(shap_explainer, X_test, shap_values)
86
87 # ===== 8. Model Deployment =====
88 # Save the trained model
89 joblib.dump(rf_model, 'model.pkl')
90
91 # Load the saved model
92 loaded_model = joblib.load('model.pkl')
93
94 # Predict on new data
95 new_data = pd.read_csv('new_data.csv')
96 new_data = pd.get_dummies(new_data, drop_first=True)
97 new_data_scaled = scaler.transform(new_data)
98 predictions = loaded_model.predict(new_data_scaled)
99
100 # Print predictions
101 print(predictions)

```

FIG. 1A

(EN) Provided herein are flu hemagglutinin polypeptides, including chimeric influenza virus hemagglutinin polypeptides, and flu hemagglutinin polypeptides comprising modified glycosylation sites and non-naturally glycosylation sites, compositions comprising the same, vaccines comprising the same and methods of their use.

FORM 2

THE PATENTS ACT, 1970
(39 of 1970)
AND
THE PATENTS RULES, 2003

**COMPLETE
SPECIFICATION**

(See Section 10; rule 13)

TITLE OF THE INVENTION

“INFLUENZA VIRUS VACCINES AND USES THEREOF”

APPLICANT

MOUNT SINAI SCHOOL OF MEDICINE
of One Gustave L. Levy Place, New York, New York 10029, USA; Nationality: USA

The following specification particularly describes
the invention and the manner in which
it is to be performed

WHAT IS CLAIMED IS:

1. A chimeric influenza virus hemagglutinin (HA) polypeptide comprising an HA stem domain and an HA globular head domain, wherein the HA globular head domain is heterologous to the HA stem domain.
2. A chimeric influenza virus hemagglutinin (HA) polypeptide comprising an HA stem domain of a seasonal influenza virus strain and an HA globular head domain of a heterologous influenza virus strain.
3. The chimeric influenza virus HA polypeptide of claim 1 or 2, wherein the HA stem domain maintains the cysteine residues designated Ap and Aq in Figure 1.
4. The chimeric influenza virus HA polypeptide of claim 1 or 2, wherein the HA stem domain is the HA stem domain of an influenza virus of subtype H1.
5. The chimeric influenza virus HA polypeptide of claim 1 or 2, wherein the HA stem domain is the HA stem domain of an influenza virus of subtype H3.
6. The chimeric influenza virus HA polypeptide of claim 4, wherein the HA globular head domain is the HA globular head domain of an influenza virus of subtype H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, H16, or H17.
7. The chimeric influenza virus HA polypeptide of claim 5, wherein the HA globular head domain is the HA globular head domain of an influenza virus of subtype H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, H16, or H17.
8. The chimeric influenza virus HA polypeptide of claim 6, wherein the HA globular head domain is the HA globular head domain of an influenza virus of subtype H5, H6, H8, H9, H11, H12, H13, or H16.
9. The chimeric influenza virus HA polypeptide of claim 7, wherein the HA globular head domain is the HA globular head domain of an influenza virus of subtype H4, H7, H10, H14, or H15.

10. A nucleic acid encoding the polypeptide of any of claims 1 to 9.
11. A cell expressing the nucleic acid of claim 10.
12. A virus comprising a genome engineered to express the nucleic acid of claim 10.
13. A virus comprising the polypeptide of any one of claims 1 to 9.
14. The virus of claim 12, wherein the virus is an influenza virus.
15. The virus of claim 13, wherein the virus is an influenza virus.
16. The virus of claim 14, which is inactivated or split.
17. The virus of claim 15, which is inactivated or split.
18. A virus-like particle comprising the polypeptide of any one of claims 1 to 9.
19. An immunogenic composition comprising the polypeptide of any one of claims 1 to 9.
20. An immunogenic composition comprising the virus of claim 12.
21. An immunogenic composition comprising the virus of claim 13.
22. An immunogenic composition comprising the virus of claim 14.
23. An immunogenic composition comprising the virus of claim 15.
24. An immunogenic composition comprising the virus-like particle of claim 18.
25. A method of immunizing a subject comprising administering to the subject an effective amount of the immunogenic composition of claim 19.

26. A method of immunizing a subject comprising administering to the subject an effective amount of the immunogenic composition of claim 20.

27. A method of immunizing a subject comprising administering to the subject an effective amount of the immunogenic composition of claim 21.

28. A method of immunizing a subject comprising administering to the subject an effective amount of the immunogenic composition of claim 22.

29. A method of immunizing a subject comprising administering to the subject an effective amount of the immunogenic composition of claim 23.

30. A method of immunizing a subject comprising administering to the subject an effective amount of the immunogenic composition of claim 24.

31. The method of claim 25, wherein the subject is a human.

32. The method of claim 26, wherein the subject is a human.

33. The method of claim 27, wherein the subject is a human.

34. The method of claim 28, wherein the subject is a human.

35. The method of claim 29, wherein the subject is a human.

36. The method of claim 30, wherein the subject is a human.

37. The method of claim 25, wherein the immunogenic composition is administered intramuscularly or intranasally to the subject.

38. A method of preventing an influenza virus disease comprising administering to a subject an effective amount of the immunogenic composition of claim 25.

39. A method of treating an influenza virus infection or an influenza virus disease comprising administering to a subject an effective amount of the immunogenic composition of claim 25.

40. A method of preventing an influenza virus disease comprising administering to a subject an effective amount of the immunogenic composition of claim 25.

41. A method of preventing an influenza virus disease comprising administering to a subject an effective amount of the immunogenic composition of claim 25.

42. The method of claim 26, wherein the immunogenic composition is administered intramuscularly or intranasally to the subject.

43. A method of preventing an influenza virus disease comprising administering to a subject an effective amount of the immunogenic composition of claim 26.

44. A method of treating an influenza virus infection or an influenza virus disease comprising administering to a subject an effective amount of the immunogenic composition of claim 26.

45. A method of preventing an influenza virus disease comprising administering to a subject an effective amount of the immunogenic composition of claim 26.

46. A method of preventing an influenza virus disease comprising administering to a subject an effective amount of the immunogenic composition of claim 26.

47. A method of preventing an influenza virus disease or infection in a subject, comprising administering to said subject an influenza virus hemagglutinin polypeptide to which the subject is naive.

48. The method of claim 47, wherein the influenza virus hemagglutinin polypeptide is from an influenza virus of subtype H2, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, H16, and/or H17.

49. The method of claim 48, comprising administering to said subject a second influenza virus hemagglutinin polypeptide to which the subject is naive, wherein the second influenza virus hemagglutinin polypeptide is from an influenza virus of subtype H2, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, H16, and/or H17, and wherein the influenza virus hemagglutinin polypeptides of the first and second administrations are from different influenza virus subtypes.

50. The method of claim 49, comprising administering to said subject a third influenza virus hemagglutinin polypeptide to which the subject is naive, wherein the third influenza virus hemagglutinin polypeptide is from an influenza virus of subtype H2, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, H16, and/or H17, and wherein the influenza virus hemagglutinin polypeptides of the first, second, and third administrations are from different influenza virus subtypes.

51. A method of preventing an influenza virus disease or infection in a subject, comprising administering to said subject an influenza virus, wherein said influenza virus comprises a hemagglutinin polypeptide to which the subject is naive.

52. The method of claim 51, wherein the influenza virus hemagglutinin polypeptide is from an influenza virus of subtype H2, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, H16, and/or H17.

53. The method of claim 52, comprising administering to said subject a second influenza virus, wherein said second influenza virus comprises a hemagglutinin polypeptide to which the subject is naive; wherein the hemagglutinin polypeptide of the second influenza virus is from an influenza virus of subtype H2, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13, H14, H15, H16, and/or H17; and wherein the influenza virus hemagglutinin polypeptides of the first and second administrations are from different influenza virus subtypes.

54. The method of claim 53, comprising administering to said subject a third influenza virus, wherein said third influenza virus comprises a hemagglutinin polypeptide to which the subject is naive; wherein the hemagglutinin polypeptide of the third influenza virus is from an influenza virus of subtype H2, H4, H5, H6, H7, H8, H9, H10, H11, H12, H13,

H14, H15, H16, and/or H17; and wherein the influenza virus hemagglutinin polypeptides of the first, second, and third administrations are from different influenza virus subtypes.

55. A chimeric influenza virus hemagglutinin (HA) polypeptide comprising an HA stem domain and an HA globular head domain, wherein the HA globular head domain is heterologous to the HA stem domain, wherein the HA stem domain comprises at least one modified glycosylation site, wherein the modified glycosylation site comprises a modification of a naturally occurring glycosylation site having an amino acid sequence Asn-Xaa-Ser/Thr/Cys, and wherein the modification disrupts the ability of a glycan to attach to the modified glycosylation site, and wherein Xaa is any amino acid.

56. The chimeric influenza virus HA polypeptide of claim 55, wherein the modification comprises one or more amino acid substitutions of the naturally occurring glycosylation site.

57. The chimeric influenza virus HA polypeptide of claim 55, wherein the modified glycosylation site is located at amino acid positions selected from the group consisting of amino acid positions 20-22 21-23, 33-35, 46-48, 289-291, 290-292, 296-298 and 481 483, according to the H3 numbering system, wherein the HA stem domain is an HA stem domain from an influenza virus of subtype H1, H2, H5, H6, H8, H9, H11, H12, H13, and H16.

58. The chimeric influenza virus HA polypeptide of claim 55, wherein the modified glycosylation site is located at amino acid positions selected from the group consisting of amino acid positions 8-10, 22-24, 38-40, 46-48, 296-298, 410-412, and 481-483, according to the H3 numbering system, wherein the HA stem domain is an HA stem domain from an influenza virus of subtype H3, H4, H7, H10, H14, or H15.

59. The chimeric influenza virus HA polypeptide of any one of claims 55-58, wherein the HA globular head domain further comprises one or more non-naturally occurring glycosylation sites having an amino acid sequence Asn-Xaa-Ser/Thr/Cys, wherein Xaa is any amino acid.

60. The chimeric influenza virus hemagglutinin polypeptide of claim 59, wherein the HA globular head domain is the HA globular head domain of an influenza virus H1 subtype and wherein the non-naturally occurring glycosylation site is located in an Sa, Sb, Ca, or Cb antigenic site; or wherein the HA globular head domain is an influenza virus H3 subtype and wherein the non-naturally occurring glycosylation site is located in an A, B, C, or D antigenic site.

61. The chimeric influenza virus hemagglutinin (HA) polypeptide of claim 59, wherein the non-naturally occurring glycosylation site is at hemagglutinin amino acid positions 59-61, 129-131, 158-160 or 165-167, according to H3 numbering.

62. A chimeric influenza virus hemagglutinin (HA) polypeptide comprising an HA stem domain and an HA globular head domain, wherein the HA globular head domain is heterologous to the HA stem domain, and wherein the HA globular head domain comprises one or more non-naturally occurring glycosylation sites having an amino acid sequence Asn-Xaa-Ser/Thr/Cys, and wherein Xaa is any amino acid.

63. A non-chimeric influenza virus hemagglutinin (HA) polypeptide comprising an HA stem domain and an HA globular head domain, wherein the HA globular head domain is homologous to the HA stem domain, wherein the HA stem domain comprises one or more modified glycosylation sites, wherein the modified glycosylation site comprises a modification of a naturally occurring glycosylation site having an amino acid sequence Asn-Xaa-Ser/Thr/Cys, and wherein the modification disrupts the ability of a glycan to attach to the modified glycosylation site, wherein Xaa is any amino acid.

64. The non-chimeric influenza virus HA polypeptide of claim 63, wherein the HA globular head domain further comprises one or more non-naturally occurring glycosylation site having an amino acid sequence Asn-Xaa-Ser/Thr/Cys, wherein Xaa is any amino acid.

65. The non-chimeric influenza virus hemagglutinin polypeptide of claim 64, wherein the HA globular head domain is the HA globular head domain of an influenza virus H1 subtype and wherein the non-naturally occurring glycosylation site is located in an Sa, Sb, Ca, or Cb antigenic site; or wherein the HA globular head domain is an influenza virus H3

subtype and wherein the non-naturally occurring glycosylation site is located in an A, B, C, or D antigenic site.

66. The non-chimeric influenza virus hemagglutinin (HA) polypeptide of claim 64 or 65, wherein the non-naturally occurring glycosylation site is at hemagglutinin amino acid position 59-61, 81-83, 129-131, 143-145, 158-160 and/or 165-167, 170-172, 187-189, 193-195, 197-199, 208-210, according to H3 numbering.

67. An influenza virus hemagglutinin (HA) stem domain polypeptide comprising:
a. an influenza hemagglutinin HA1 domain that comprises an HA1 N-terminal stem segment covalently linked to a linker of 1 to 50 heterologous residues that is in turn covalently linked to an HA1 C-terminal short stem segment; said HA1 domain in tertiary or quaternary association with

b. an influenza hemagglutinin HA2 domain,

wherein the influenza virus HA stem domain polypeptide domain further comprises one or more modified glycosylation sites, wherein the modified glycosylation site comprises a modification of a naturally occurring glycosylation site having an amino acid sequence Asn-Xaa-Ser/Thr/Cys, where the modification disrupts the ability of a glycan to attach to the modified glycosylation site, and wherein Xaa is any amino acid.

68. An influenza virus hemagglutinin (HA) stem domain polypeptide comprising:
a. an influenza hemagglutinin HA1 domain that comprises an HA1 N-terminal long stem segment covalently linked to a linker of 1 to 50 heterologous residues that is in turn covalently linked to an HA1 C-terminal long stem segment; said HA1 domain in tertiary or quaternary association with

b. an influenza hemagglutinin HA2 domain,

wherein the influenza virus HA stem domain polypeptide domain further comprises one or more modified glycosylation site, wherein the modified glycosylation site comprises a modification of a naturally occurring glycosylation site having an amino acid sequence Asn-Xaa-Ser/Thr/Cys, where the modification disrupts the ability of a glycan to attach to the modified glycosylation site, and wherein Xaa is any amino acid.

69. An influenza virus hemagglutinin (HA) stem domain polypeptide comprising:

a. an influenza hemagglutinin HA1 domain that comprises an HA1 N-terminal stem segment covalently linked to a linker of 1 to 50 heterologous residues that is in turn covalently linked to an HA1 C-terminal stem segment; said HA1 domain in tertiary or quaternary association with

b. an influenza hemagglutinin HA2 domain,

wherein the influenza virus HA stem domain polypeptide domain further comprises one or more modified glycosylation site, wherein the modified glycosylation site comprises a modification of a naturally occurring glycosylation site having an amino acid sequence Asn-Xaa-Ser/Thr/Cys, where the modification disrupts the ability of a glycan to attach to the modified glycosylation site, and wherein Xaa is any amino acid.

70. An influenza virus hemagglutinin (HA) stem domain polypeptide comprising:

a. an influenza hemagglutinin HA1 domain that comprises, linked in the following order: an HA1 N-terminal stem segment, a first linker of 1 to 50 heterologous residues, an HA1 intermediate stem segment, a second linker of 1 to 50 heterologous residues and an HA1 C-terminal stem segment; said HA1 domain in tertiary or quaternary association with

b. an influenza hemagglutinin HA2 domain,

wherein the influenza virus hemagglutinin (HA) stem domain polypeptide domain further comprises one or more modified glycosylation site, wherein the modified glycosylation site comprises a modification of a naturally occurring glycosylation site having an amino acid sequence Asn-Xaa-Ser/Thr/Cys, where the modification disrupts the ability of a glycan to attach to the modified glycosylation site, and wherein Xaa is any amino acid.

71. The influenza virus hemagglutinin (HA) stem domain polypeptide of any one of claims 68 to 71, wherein the modified glycosylation site is located at amino acid positions selected from the group consisting of amino acid positions 20-22, 21-23, 33-35, 46-48, 289-291, 290-292, 296-298 and 481-483, according to the H3 numbering system, wherein the HA stem domain is an HA stem domain from an influenza virus of subtype H1, H2, H5, H6, H8, H9, H11, H12, H13, and H16.

72. The influenza virus hemagglutinin (HA) stem domain polypeptide of any one of claims 68 to 71, wherein the modified glycosylation site is located at amino acid positions selected from the group consisting of amino acid positions 8-10, 22-24, 38-40, 46-48, 296-

298, 410-412, and 481-483, according to the H3 numbering system, wherein the HA stem domain is an HA stem domain from an influenza virus of subtype H3, H4, H7, H10, H14, H15.

73. The HA polypeptide of any one of claims 1-9 or 55-72, wherein said polypeptide is soluble.

74. A nucleic acid encoding the polypeptide of any of claims 55-62.

75. A nucleic acid encoding the polypeptide of any of claims 63-66.

76. A cell expressing the nucleic acid of claim 74.

77. A cell expressing the nucleic acid of claim 75.

78. A virus comprising the chimeric influenza virus hemagglutinin (HA) polypeptide of any one of claims 55-62.

79. A virus comprising the chimeric influenza virus hemagglutinin (HA) polypeptide of any one of claims 63-66.

80. An immunogenic composition comprising the chimeric influenza virus hemagglutinin (HA) polypeptide of any one of claims 55-62.

81. An immunogenic composition comprising the chimeric influenza virus hemagglutinin (HA) polypeptide of any one of claims 63-66.

82. A method of immunizing a subject comprising administering to the subject an effective amount of the immunogenic composition of claim 80.

83. A method of immunizing a subject comprising administering to the subject an effective amount of the immunogenic composition of claim 81.

84. The method of claim 82, wherein the subject is human.

85. The method of claim 83, wherein the subject is human.

Dated this 19 day of March 2014

Arindam Paul
REG.NO:IN/PA-174
of Depenning & Depenning
Agent for the Applicants