Text Classification : 20 Newsgroups

```
# Text Classification : 20 Newsgroups
# Author : Yassine Ghouzam
# Date : 28/03/2017
# Version : 1
# Kernel version : 1
# Output version : 1
# Input : 20 Newsgroups
# Type : Text Classification
# Category : Supervised
# Number of Instances : 18846
# Number of Attributes : 1
# Attribute Information : text, category
# Source : http://qwone.com/~jason/20Newsgroups/
# Description : 20 Newsgroups is a collection of approximately 20,000 newsgroup documents, partitioned (nearly) evenly across 20 different newsgroups.
# Relevant Information : 20 Newsgroups is a collection of approximately 20,000 newsgroup documents, partitioned (nearly) evenly across 20 different newsgroups. To the best of our knowledge, it was originally collected by Ken Lang, probably for his Newsweeder: Learning to filter netnews paper, though he does not explicitly mention this collection. The 20 newsgroups collection has become a popular data set for experiments in text applications of machine learning techniques, such as text classification and text clustering.
```

```
# Import of libraries
import numpy as np
# Linear algebra
import pandas as pd
# Data processing, CSV file I/O (e.g. pd.read_csv)
from sklearn.cross_validation import train_test_split,cross_val_score
from sklearn.feature_extraction.text import CountVectorizer,TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB,BernoulliNB
from sklearn.metrics import classification_report,confusion_matrix,accuracy_score
from sklearn.pipeline import Pipeline
```

```
# Import of dataset
df = pd.read_csv('../input/newsgroups20.zip/newsgroups20/')
df.shape
# Displays the number of rows and columns in the dataset (rows, columns)
df.columns.tolist() # Gives a list of column names)
df.head(5)
# Displays the first 5 rows
df.tail(5)
# Displays the last 5 rows
df.dtypes
# Displays the data types of each column
df.nunique(axis=0)
# Displays the number of unique values in each column
df.nunique(axis=1)
df.shape[0] - df.nunique(axis=1)
# Displays the number of duplicate records
df[df.duplicated(keep='first')]
# Displays all duplicate records
df.isnull().sum()
# Displays the number of missing values in each column
df.isnull().sum(axis=1)
# Displays the number of missing values in each row
df.describe(include=['object'])
# Displays the statistical summary of the dataset
```

```
# Displays the count, unique, top, and freq of the dataset
df.groupby('category').text.count()
# Displays the count of each category
```

```
# Displays the count of each category in the dataset
df.groupby('category').text.count().plot.bar(ylim=0)
# Displays the bar plot of the count of each category
```

```
# Displays the bar plot of the count of each category
df['text_length'] = df['text'].apply(lambda x: len(x.split(' '),dtype=np.int64))
# Adds a column to the dataset that contains the length of each text
df.groupby('category').text_length.sum().plot.bar(ylim=0)
# Displays the bar plot of the sum of text length for each category
```

```
# Displays the bar plot of the sum of text length for each category
df.groupby('category').text_length.mean().plot.bar(ylim=0)
# Displays the bar plot of the mean of text length for each category
```

```
# Displays the bar plot of the mean of text length for each category
X = df['text']
# Assigns the text column to X
y = df['category']
# Assigns the category column to y
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=1)
# Splits the dataset into training and testing sets
vect = CountVectorizer()
# Initializes the CountVectorizer
X_train_dtm = vect.fit_transform(X_train)
# Transforms the training set into a document-term matrix
X_test_dtm = vect.transform(X_test)
# Transforms the testing set into a document-term matrix
print(X_train_dtm.shape)
# Displays the shape of the training set
print(X_test_dtm.shape)
# Displays the shape of the testing set
nb = MultinomialNB()
# Initializes the Multinomial Naive Bayes classifier
nb.fit(X_train_dtm, y_train)
# Fits the classifier to the training set
y_pred_class = nb.predict(X_test_dtm)
# Predicts the category of the testing set
print(metrics.accuracy_score(y_test, y_pred_class))
# Displays the accuracy score of the classifier
print(metrics.confusion_matrix(y_test, y_pred_class))
# Displays the confusion matrix of the classifier
print(metrics.classification_report(y_test, y_pred_class))
# Displays the classification report of the classifier
```

```
# Displays the classification report of the classifier
nb = Pipeline([('vect', CountVectorizer()),
               ('tfidf', TfidfTransformer()),
               ('nb', MultinomialNB())])
# Initializes the pipeline with the CountVectorizer, TfidfTransformer and MultinomialNB classifier
nb.fit(X_train, y_train)
# Fits the pipeline to the training set
y_pred_class = nb.predict(X_test)
# Predicts the category of the testing set
print(metrics.accuracy_score(y_test, y_pred_class))
# Displays the accuracy score of the classifier
print(metrics.confusion_matrix(y_test, y_pred_class))
# Displays the confusion matrix of the classifier
print(metrics.classification_report(y_test, y_pred_class))
# Displays the classification report of the classifier
```

Regulation of [redacted] in the CNS/PNS by miRNA-21-5p in ischemic stroke

Introduction Ischemic stroke is a cerebrovascular disease caused by cerebral ischemia, hypoxia, and necrosis, which is characterized by high incidence, high disability, high recurrence, high mortality, and high economic burden. Ischemic stroke accounts for approximately 87% of all strokes, and the pathogenesis of ischemic stroke is complex, involving multiple pathophysiological processes such as excitotoxicity, oxidative stress, inflammation, and apoptosis. Although thrombolysis and thrombectomy have been widely used in the treatment of acute ischemic stroke, the narrow therapeutic window limits their clinical application. Therefore, exploring new therapeutic targets and strategies is of great clinical significance. MicroRNA-21-5p (miR-21-5p) is a highly conserved microRNA that plays a key regulatory role in various physiological and pathological processes. Studies have shown that miR-21-5p is abnormally expressed in ischemic stroke and may regulate neuroinflammation, blood-brain barrier integrity, and neuronal apoptosis through multiple signaling pathways. This review systematically summarizes the expression, regulatory mechanisms, and potential clinical applications of miR-21-5p in ischemic stroke, aiming to provide new theoretical basis and research directions for the diagnosis and treatment of ischemic stroke.

Methods We searched PubMed, Web of Science, and Chinese National Knowledge Infrastructure databases using the keywords "miR-21-5p", "ischemic stroke", "cerebral ischemia", "neuroinflammation", "blood-brain barrier", and "neuronal apoptosis". The search was conducted from January 2010 to December 2024. Inclusion criteria were: (1) original research on miR-21-5p in ischemic stroke; (2) clinical or animal experimental studies; (3) English or Chinese language. Exclusion criteria were: (1) reviews, meta-analyses, and conference abstracts; (2) studies on other types of stroke; (3) studies on other miR-21 isoforms. A total of 1,247 articles were initially retrieved, and after screening, 89 articles were included in this review.

Results miR-21-5p is widely expressed in the brain and peripheral blood. In ischemic stroke, miR-21-5p expression is dynamically changed, showing early elevation and subsequent gradual decline. miR-21-5p regulates neuroinflammation by targeting PDCD4, PTEN, and TLR4/NF-κB signaling pathways, thereby inhibiting the release of pro-inflammatory cytokines and promoting anti-inflammatory responses. It also protects the blood-brain barrier by regulating tight junction proteins and matrix metalloproteinases. miR-21-5p inhibits neuronal apoptosis through the PTEN/Akt/mTOR pathway and promotes neurogenesis and angiogenesis. In addition, miR-21-5p may serve as a biomarker for early diagnosis and prognosis assessment of ischemic stroke, and its circulating levels are correlated with stroke severity and patient prognosis.

Conclusion miR-21-5p plays a multifaceted role in the pathophysiology of ischemic stroke, including regulating neuroinflammation, protecting the blood-brain barrier, inhibiting neuronal apoptosis, and promoting neurogenesis and angiogenesis. It has potential as a biomarker and therapeutic target for ischemic stroke. However, most studies are still at the preclinical stage, and further clinical trials are needed to verify its safety and efficacy in humans. Future research should also explore the upstream regulatory mechanisms of miR-21-5p and its interactions with other non-coding RNAs.

Introduction Ischemic stroke is a cerebrovascular disease caused by cerebral ischemia, hypoxia, and necrosis, which is characterized by high incidence, high disability, high recurrence, high mortality, and high economic burden. Ischemic stroke accounts for approximately 87% of all strokes, and the pathogenesis of ischemic stroke is complex, involving multiple pathophysiological processes such as excitotoxicity, oxidative stress, inflammation, and apoptosis. Although thrombolysis and thrombectomy have been widely used in the treatment of acute ischemic stroke, the narrow therapeutic window limits their clinical application. Therefore, exploring new therapeutic targets and strategies is of great clinical significance. MicroRNA-21-5p (miR-21-5p) is a highly conserved microRNA that plays a key regulatory role in various physiological and pathological processes. Studies have shown that miR-21-5p is abnormally expressed in ischemic stroke and may regulate neuroinflammation, blood-brain barrier integrity, and neuronal apoptosis through multiple signaling pathways. This review systematically summarizes the expression, regulatory mechanisms, and potential clinical applications of miR-21-5p in ischemic stroke, aiming to provide new theoretical basis and research directions for the diagnosis and treatment of ischemic stroke.

Methods We searched PubMed, Web of Science, and Chinese National Knowledge Infrastructure databases using the keywords "miR-21-5p", "ischemic stroke", "cerebral ischemia", "neuroinflammation", "blood-brain barrier", and "neuronal apoptosis". The search was conducted from January 2010 to December 2024. Inclusion criteria were: (1) original research on miR-21-5p in ischemic stroke; (2) clinical or animal experimental studies; (3) English or Chinese language. Exclusion criteria were: (1) reviews, meta-analyses, and conference abstracts; (2) studies on other types of stroke; (3) studies on other miR-21 isoforms. A total of 1,247 articles were initially retrieved, and after screening, 89 articles were included in this review.

Results miR-21-5p is widely expressed in the brain and peripheral blood. In ischemic stroke, miR-21-5p expression is dynamically changed, showing early elevation and subsequent gradual decline. miR-21-5p regulates neuroinflammation by targeting PDCD4, PTEN, and TLR4/NF-κB signaling pathways, thereby inhibiting the release of pro-inflammatory cytokines and promoting anti-inflammatory responses. It also protects the blood-brain barrier by regulating tight junction proteins and matrix metalloproteinases. miR-21-5p inhibits neuronal apoptosis through the PTEN/Akt/mTOR pathway and promotes neurogenesis and angiogenesis. In addition, miR-21-5p may serve as a biomarker for early diagnosis and prognosis assessment of ischemic stroke, and its circulating levels are correlated with stroke severity and patient prognosis.

Conclusion miR-21-5p plays a multifaceted role in the pathophysiology of ischemic stroke, including regulating neuroinflammation, protecting the blood-brain barrier, inhibiting neuronal apoptosis, and promoting neurogenesis and angiogenesis. It has potential as a biomarker and therapeutic target for ischemic stroke. However, most studies are still at the preclinical stage, and further clinical trials are needed to verify its safety and efficacy in humans. Future research should also explore the upstream regulatory mechanisms of miR-21-5p and its interactions with other non-coding RNAs.

Transcriptionists, Transcriptionists, Transcriptionists,

{"Ahmadabad": 1, "Bhopal": 1, "Chennai": 1, "Ghaziabad": 1, "Indore": 1, "Jaipur": 1, "Kanpur": 1, "Lucknow": 1, "Nagpur": 1, "Patna": 1, "Pune": 1, "Surat": 1, "Vadodara": 1, "Varanasi": 1}

```
# Analyze which cities appear in Indian Cities.csv
cities = df['City'].value_counts()
print("Unique cities:", len(cities))
print(cities.head(20))
print("\n", cities.tail(20))
```

[illegible]

```
[ 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000
```

Accuracy Comparison :

```
# Calculate accuracy for Random Forest
rf_accuracy = accuracy_score(y_test, rf_predictions)
print("Random Forest Accuracy:", rf_accuracy)

# Calculate accuracy for Decision Tree
dt_accuracy = accuracy_score(y_test, dt_predictions)
print("Decision Tree Accuracy:", dt_accuracy)
```