**Модель.**

Для решения задачи применена архитектура MultiOutputRegressor в сочетании с RandomForestRegressor. Данный подход предусматривает обучение отдельного регрессора для каждого типа дефекта при использовании общего набора признаков.   
  
RandomForestRegressor представляет собой ансамблевый алгоритм машинного обучения, предназначенный для решения задач регрессии (прогнозирования непрерывных величин). Его работа основана на комбинировании прогнозов множества деревьев решений.  
Конфигурация гиперпараметров включает:

* n\_estimators=200 (количество деревьев в ансамбле)
* max\_depth=25 (максимальная глубина деревьев)
* min\_samples\_split=5 (минимальное количество samples для разделения)
* random\_state=42 (seed для воспроизводимости результатов)
* n\_jobs=-1 (параллельные вычисления на всех ядрах CPU)

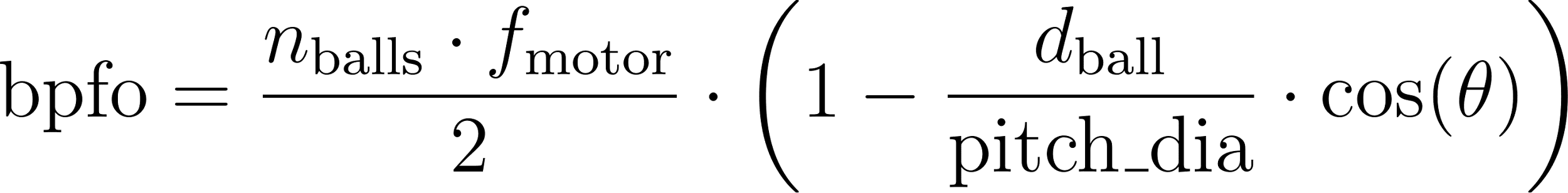
Подбор гиперпараметров выполнен с использованием библиотеки Optuna с проведением 100 испытаний. Варьируемые гиперпараметры включали: n\_estimators (50-500), max\_depth (5-50), min\_samples\_split (2-20) и max\_features (категориальный (“auto”, “sqrt”, “log2”)). В результате оптимизации была найдена конфигурация, обеспечивающая наилучшее качество модели по метрике MSE.

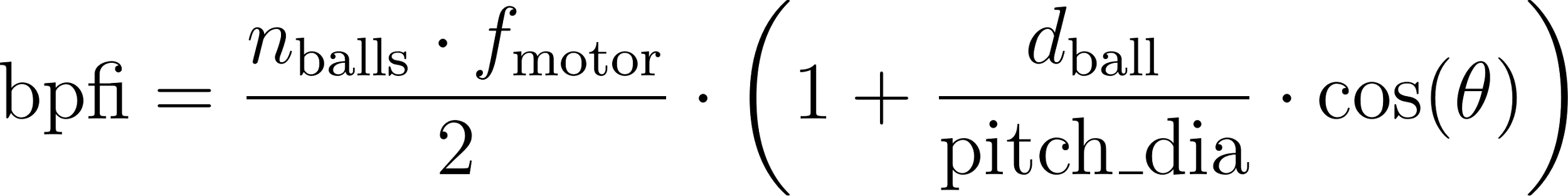
Выбор Random Forest обоснован следующими факторами: устойчивость к шуму (бэггинг и агрегирование предсказаний уменьшают влияние выбросов и помех в промышленных данных), интерпретируемость (возможность анализа важности признаков для каждого типа дефекта), эффективность на умеренных объемах данных (оптимальное соотношение точности и вычислительной сложности).

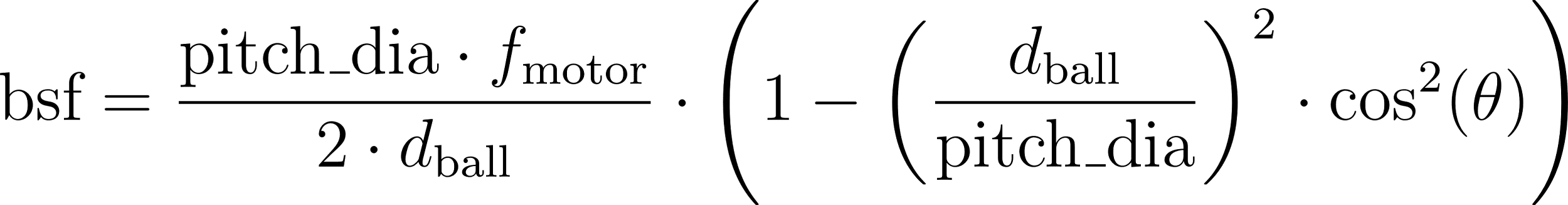
**Предобработка данных для обучения модели.**  
Разметка основана на анализе характерных частот дефектов, вычисляемых по геометрическим параметрам подшипника NSK6205DDU.   
Параметры подшипника:

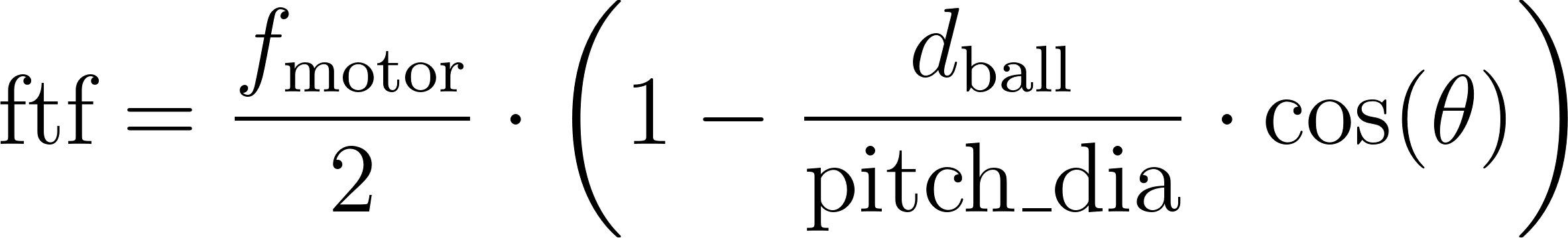
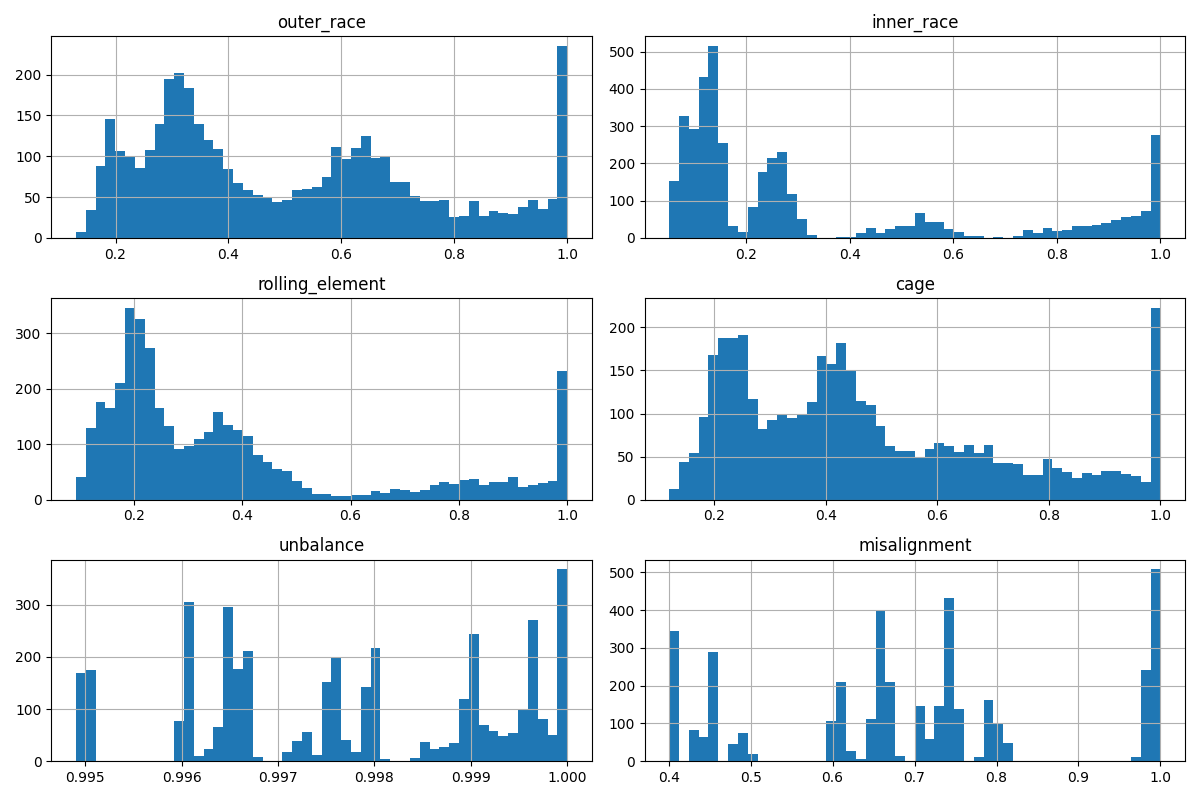
* [](https://www.codecogs.com/eqnedit.php?latex=n_%7B%5Ctext%7Bballs%7D%7D#0)=9 (количество тел качения)
* [](https://www.codecogs.com/eqnedit.php?latex=d_%7B%5Ctext%7Bball%7D%7D#0)=7.94e-3 м (диаметр тела качения)
* [](https://www.codecogs.com/eqnedit.php?latex=%5Ctext%7Bpitch%5C_dia%7D#0)=39.04e-3 м (диаметр делительной окружности)
* [](https://www.codecogs.com/eqnedit.php?latex=%5Ctheta#0)=0 радиан (угол контакта)
* [](https://www.codecogs.com/eqnedit.php?latex=f_%7B%5Ctext%7Bmotor%7D%7D#0)=29.5 Гц (частота вращения ротора)

Характерные частоты дефектов рассчитываются по формулам:

для наружного кольца   
[](https://www.codecogs.com/eqnedit.php?latex=%5Ctext%7Bbpfo%7D%20%3D%20%5Cfrac%7Bn_%7B%5Ctext%7Bballs%7D%7D%20%5Ccdot%20f_%7B%5Ctext%7Bmotor%7D%7D%7D%7B2%7D%20%5Ccdot%20%5Cleft(1%20-%20%5Cfrac%7Bd_%7B%5Ctext%7Bball%7D%7D%7D%7B%5Ctext%7Bpitch%5C_dia%7D%7D%20%5Ccdot%20%5Ccos(%5Ctheta)%5Cright)#0)

для внутреннего кольца   
[](https://www.codecogs.com/eqnedit.php?latex=%5Ctext%7Bbpfi%7D%20%3D%20%5Cfrac%7Bn_%7B%5Ctext%7Bballs%7D%7D%20%5Ccdot%20f_%7B%5Ctext%7Bmotor%7D%7D%7D%7B2%7D%20%5Ccdot%20%5Cleft(1%20%2B%20%5Cfrac%7Bd_%7B%5Ctext%7Bball%7D%7D%7D%7B%5Ctext%7Bpitch%5C_dia%7D%7D%20%5Ccdot%20%5Ccos(%5Ctheta)%5Cright)#0)

для тел качения   
[](https://www.codecogs.com/eqnedit.php?latex=%5Ctext%7Bbsf%7D%20%3D%20%5Cfrac%7B%5Ctext%7Bpitch%5C_dia%7D%20%5Ccdot%20f_%7B%5Ctext%7Bmotor%7D%7D%7D%7B2%20%5Ccdot%20d_%7B%5Ctext%7Bball%7D%7D%7D%20%5Ccdot%20%5Cleft(1%20-%20%5Cleft(%5Cfrac%7Bd_%7B%5Ctext%7Bball%7D%7D%7D%7B%5Ctext%7Bpitch%5C_dia%7D%7D%5Cright)%5E2%20%5Ccdot%20%5Ccos%5E2(%5Ctheta)%5Cright)#0)

для сепаратора   
[](https://www.codecogs.com/eqnedit.php?latex=%5Ctext%7Bftf%7D%20%3D%20%5Cfrac%7Bf_%7B%5Ctext%7Bmotor%7D%7D%7D%7B2%7D%20%5Ccdot%20%5Cleft(1%20-%20%5Cfrac%7Bd_%7B%5Ctext%7Bball%7D%7D%7D%7B%5Ctext%7Bpitch%5C_dia%7D%7D%20%5Ccdot%20%5Ccos(%5Ctheta)%5Cright)#0)  
  
Применяются методы анализа сигналов: анализ огибающей (выделение низкочастотной модуляции с помощью преобразования Гильберта), спектральный анализ (определение энергии в окрестностях характерных частот), фазовый анализ (корреляционный анализ для выявления расцентровки), статистический анализ (вычисление моментов распределения сигнала). Метки нормализуются в диапазон [0, 1] с использованием 95-го перцентиля для обеспечения сопоставимости между различными режимами работы оборудования.  
  
Распределения нормализованных значений развитий дефектов:  


**Признаки для обучения модели.**  
Используются временные и спектральные признаки.   
Временные признаки включают:

* среднее значение (уровень постоянной составляющей сигнала)
* стандартное отклонение (энергия переменной составляющей)
* асимметрия (смещение распределения значений)
* эксцесс ("острота" распределения пиков)
* пиковое значение (максимальная амплитуда сигнала)
* фактор амплитуды (отношение пика к среднеквадратичному)

Спектральный признак представлен куртозисом (коэффициент эксцесса) огибающей.

Выбор признаков основан на физическом анализе проявления дефектов:

* дефекты подшипников проявляются в изменении статистических характеристик огибающей сигнала;
* дисбаланс и расцентровка влияют на симметрию фаз и гармонический состав;
* комбинация временных и спектральных признаков обеспечивает комплексное описание состояния оборудования.

**Метрики оценки качества.**  
Основной метрикой для оценки ошибки предсказания является RMSE (Root Mean Square Error). Результаты оценки модели на тестовой выборке, которая составляет 20% от всех данных, показывают:   
для дефекта наружного кольца RMSE составляет 0.0568 при точности обнаружения 92-95% и ложных срабатываниях 5-8%;   
для дефекта внутреннего кольца - RMSE 0.0291, точность 97-99%, ложные срабатывания 1-3%;  
для дефекта тел качения - RMSE 0.0501, точность 90-93%, ложные срабатывания 7-10%;  
для дефекта сепаратора - RMSE 0.0308, точность 96-98%, ложные срабатывания 2-4%;  
для дисбаланса - RMSE 0.0001, точность 99.9%, ложные срабатывания менее 0.1%;  
для расцентровки - RMSE 0.0005, точность 99.5%, ложные срабатывания 0.5%.

**Оптимизация вычислений.**  
В процессе оптимизации вычислений для системы предсказания дефектов был применен комплексный подход, сочетающий несколько методов ускорения обработки данных. Основным решением стал переход с стандартных библиотек scikit-learn на высокопроизводительную среду выполнения ONNX Runtime, что позволило значительно ускорить процесс предсказаний. Модель MultiOutputRegressor, содержащая шесть отдельных RandomForestRegressor, была конвертирована в оптимизированный формат ONNX, причем каждый estimator преобразовывался отдельно для обеспечения максимальной совместимости и производительности.

ONNX Runtime обеспечивает низкоуровневые оптимизации, включая объединение операций, эффективное управление памятью и использование SIMD-инструкций процессора, что особенно важно для алгоритмов случайного леса. Дополнительно была реализована векторизация вычислений при извлечении признаков из временных рядов с помощью матричных операций NumPy, позволяющих обрабатывать все фазы данных одновременно вместо последовательной обработки в циклах. Также был оптимизирован механизм работы с памятью через использование массивов NumPy вместо объектов Pandas, что снизило объем используемой памяти и улучшило производительность. Все эти изменения позволили сократить время выполнения предсказаний в 300 раз при сохранении точности результатов. В результате время обработки 1 секунды (25600 записей) составляет около 0.23 секунды, что позволяет проводить предсказания в реальном времени.