Morgan_fingerprint_Embedding

December 12, 2025

[2]:
```
import pandas as pd

# Read metabolite table
pos = pd.read_csv(
    "m_MTBLS8920_LC-MS_positive_hilic_metabolite_profiling_v2_maf.tsv",
    sep="\t"
)

# Extract metabolite → smiles
metabolite_smiles = pos[["database_identifier", "smiles"]].copy()

# Drop missing SMILES
metabolite_smiles = metabolite_smiles.dropna(subset=["smiles"])

# Drop problematic metabolite you already identified
metabolite_smiles = metabolite_smiles[
    metabolite_smiles["database_identifier"] != "CHEBI:172744"
].reset_index(drop=True)

metabolite_smiles.head()
```

[2]:

	database_identifier	smiles
0	CHEBI:28619	MTBLS6894
1	CHEBI:35807	CN(C)N=O
2	CHEBI:41879	C1=CC(C=C2[C@]1([C@@]3([C@@](CC2)([C@]4([C@](C…
3	CHEBI:19092	C1CC=NC1
4	CHEBI:175520	O1C[C@](CC=2C1=CC(OC)=C(C2OC)CC=C(C)C)(C3=C4OC…

[3]:
```
import numpy as np

# sample columns
sample_cols = [
    c for c in pos.columns
    if c.startswith("Breast_") or c.startswith("control_")
]

# abundance: sample × metabolite_id
```

```
abundance = pos.set_index("database_identifier")[sample_cols].T
abundance.index.name = "sample"

# ensure numeric
abundance = abundance.apply(pd.to_numeric, errors="coerce").fillna(0)

print("Abundance shape:", abundance.shape)
```

```
Abundance shape: (57, 493)
```

[7]:
```
import numpy as np
from sklearn.preprocessing import StandardScaler

abundance_np = abundance.values.astype(float)
abundance_np = np.nan_to_num(abundance_np, nan=0.0)
abundance_np = np.log1p(abundance_np)

scaler_ab = StandardScaler()
abundance_np_scaled = scaler_ab.fit_transform(abundance_np)

abundance_scaled = pd.DataFrame(
    abundance_np_scaled,
    index=abundance.index,
    columns=abundance.columns
)
```

[8]:
```
from rdkit import Chem
from rdkit.Chem import AllChem
import numpy as np

# Morgan fingerprint parameters
FP_DIM = 2048
RADIUS = 2

morgan_dict = {}

for _, row in metabolite_smiles.iterrows():
    met_id = row["database_identifier"]
    smi = row["smiles"]

    mol = Chem.MolFromSmiles(smi)
    if mol is None:
        continue

    fp = AllChem.GetMorganFingerprintAsBitVect(
        mol,
        radius=RADIUS,
```

```
        nBits=FP_DIM
    )

    morgan_dict[met_id] = np.array(fp, dtype=np.float32)

print("Metabolites with Morgan FP:", len(morgan_dict))
```

```
[03:28:00] SMILES Parse Error: syntax error while parsing: MTBLS6894
[03:28:00] SMILES Parse Error: check for mistakes around position 1:
[03:28:00] MTBLS6894
[03:28:00] ^
[03:28:00] SMILES Parse Error: Failed parsing SMILES 'MTBLS6894' for input:
'MTBLS6894'
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
```

```
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
```

```
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
```

```
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
```

```
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
```

```
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
```

```
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] Conflicting single bond directions around double bond at index 20.
[03:28:00]   BondStereo set to STEREONONE and single bond directions set to
NONE.
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
```

```
Metabolites with Morgan FP: 444
```

```
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
```

```
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
```

```
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
```

```
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
[03:28:00] DEPRECATION WARNING: please use MorganGenerator
```

[9]:
```
valid_mets = [m for m in abundance_scaled.columns if m in morgan_dict]

abundance_final = abundance_scaled[valid_mets]

print("Metabolites used:", len(valid_mets))
```

```
Metabolites used: 444
```

[10]:
```
sample_embeddings = {}

for sample in abundance_final.index:
    emb_vec = np.zeros(FP_DIM, dtype=np.float32)

    for met in valid_mets:
        weight = abundance_final.loc[sample, met]    # normalized abundance
```

```
        emb = morgan_dict[met]                          # Morgan FP
        emb_vec += weight * emb

    sample_embeddings[sample] = emb_vec
```

[12]:
```
import pandas as pd

meta_3 = pd.read_csv("s_MTBLS8920.txt", sep="\t")

meta_clean_3 = meta_3[~meta_3["Sample Name"].str.startswith("QC")].copy()

meta_ml_3 = meta_clean_3[["Sample Name", "Factor Value[Disease]"]].copy()

meta_ml_3 = meta_ml_3.rename(columns={
    "Sample Name": "sample",
    "Factor Value[Disease]": "disease"
})

meta_ml_3["label"] = meta_ml_3["disease"].map({
    "breast cancer": 1,
    "healthy control": 0
})
```

[13]:
```
# sample × fingerprint
sample_embed_df = pd.DataFrame.from_dict(
    sample_embeddings,
    orient="index"
)
sample_embed_df.index.name = "sample"

# merge label
morgan_abundance_ML = sample_embed_df.merge(
    meta_ml_3[["sample", "label"]],
    on="sample",
    how="inner"
)

morgan_abundance_ML
```

[13]:

	sample	0	1	2	3	4	5	\
0	Breast_1	-1.577821	-63.771938	2.519383	0.0	3.617110	2.881479	
1	Breast_10	-0.462907	-131.428162	-1.773262	0.0	-0.264680	-1.044455	
2	Breast_11	-0.058059	59.832901	1.802125	0.0	4.642465	-0.187619	
3	Breast_12	-0.001556	-75.632179	-0.360007	0.0	-0.747738	-2.223782	
4	Breast_13	0.522558	-99.677444	-1.207039	0.0	-3.705719	-2.221539	

	6	7	8	…	2039	2040	2041	2042	\

```
0 -0.578136 -1.131751 -14.571426  … -1.525499  0.619345 -7.009365  0.480932
1 -3.338238 -2.205090 -12.198885  … -2.123830  0.118527 -2.376938 -1.840528
2 -1.414549 -3.552037  -1.582109  … -1.892688  1.364370 -1.210879 -2.946210
3 -0.968055 -1.586803  -8.690717  … -0.696032  1.509288 -0.605615 -3.349327
4 -2.981936 -4.464425 -19.646568  … -2.021634  0.362182 -2.012872 -2.203544

   2043      2044      2045  2046  2047  label
0   0.0 -1.841229 -0.770367   0.0   0.0      1
1   0.0  1.806760 -0.984125   0.0   0.0      1
2   0.0 -0.358407 -1.337365   0.0   0.0      1
3   0.0 -1.091852  3.288409   0.0   0.0      1
4   0.0 -0.237049  1.347519   0.0   0.0      1

[5 rows x 2050 columns]
```

[19]:
```
nonzero_cols = (morgan_abundance_ML != 0).any(axis=0)

morgan_abundance_ML = morgan_abundance_ML.loc[:, nonzero_cols]
morgan_abundance_ML
```

[19]:

	sample	0	1	2	4	5	6	\
0	Breast_1	-1.577821	-63.771938	2.519383	3.617110	2.881479	-0.578136	
1	Breast_10	-0.462907	-131.428162	-1.773262	-0.264680	-1.044455	-3.338238	
2	Breast_11	-0.058059	59.832901	1.802125	4.642465	-0.187619	-1.414549	
3	Breast_12	-0.001556	-75.632179	-0.360007	-0.747738	-2.223782	-0.968055	
4	Breast_13	0.522558	-99.677444	-1.207039	-3.705719	-2.221539	-2.981936	
5	Breast_14	-0.377028	-80.549370	0.523497	0.415671	-2.513501	-2.294846	
6	Breast_15	1.319680	69.906898	0.385082	-1.178573	0.971266	0.304147	
7	Breast_16	1.198593	30.092297	0.894240	2.330965	2.540152	0.504328	
8	Breast_17	0.323576	-11.748161	-0.108333	-0.279216	-0.562423	-1.027992	
9	Breast_18	0.161067	-21.364115	-1.977934	-1.733437	-0.819895	0.656544	
10	Breast_19	-0.129791	-93.801315	0.092601	-1.809084	-0.358170	-0.964259	
11	Breast_2	-0.355148	-102.375092	-0.353270	1.284684	1.780886	-2.087765	
12	Breast_20	0.405727	-92.887711	-3.368630	-3.529855	-0.825127	-1.571842	
13	Breast_21	0.192209	-121.396484	-2.775005	-2.959791	-1.016257	0.124716	
14	Breast_22	0.248085	39.961792	0.021861	0.960442	2.570335	0.901591	
15	Breast_23	1.445831	23.489885	5.543217	0.987101	8.335752	0.068494	
16	Breast_24	1.223900	-59.378506	-1.136493	-2.523135	-0.995864	0.473622	
17	Breast_25	-0.517244	-31.619402	-0.197218	-3.222729	3.392680	-0.375715	
18	Breast_26	0.454824	128.776215	1.935286	-3.138026	3.768036	1.291131	
19	Breast_27	0.522346	-193.339523	-4.138151	-2.873491	-2.878180	-2.191990	
20	Breast_28	0.024950	60.777237	-0.988702	1.592048	-2.017148	0.736357	
21	Breast_3	-0.005455	-71.673767	0.048156	-2.071554	-1.885629	-1.738357	
22	Breast_4	0.827522	-42.707882	-1.475319	0.554141	1.752407	1.004571	
23	Breast_5	-0.097325	-108.588547	-0.335895	0.136871	-2.820122	1.358515	
24	Breast_7	0.439780	-17.035620	0.657089	2.373134	-2.140277	-0.660395	
25	Breast_8	0.498623	-110.746346	0.487303	-3.546439	-0.097050	-2.070427	

```
26    Breast_9  0.506481  -59.494576 -0.078015 -0.242320  1.455656  2.479789
27   control_1  0.344726   17.284214 -0.708481 -1.345473 -1.236349  0.575228
28  control_10  0.159251   40.132339  0.330227 -0.200172  0.905295  0.152785
29  control_11 -0.537176  106.829453  0.822135  0.122445  1.951240  2.054050
30  control_12  0.628927    2.086747 -0.192534 -1.490362  2.925764 -0.786773
31  control_13 -0.228658  155.230728  2.124144  1.717009  2.589468  1.904789
32  control_14  0.621906   10.887565 -1.520953 -1.154276 -1.938136  0.908089
33  control_15  0.145974   77.623772 -1.307395 -0.119290 -3.520546  1.229993
34  control_16 -0.960635   50.664139  1.293004  2.671135  2.932807  1.062628
35  control_17 -0.336467   92.861343  0.398912  0.603929 -0.457083  1.015026
36  control_18 -6.245316   78.397560  1.716001  2.673788  0.110686  0.654185
37  control_19 -0.229567  -47.549133 -0.881512 -3.440107  0.084481 -1.741984
38   control_2 -0.264012   31.794476  1.933762  4.124017  0.589679  1.498574
39  control_20  0.302372   28.694813 -0.406864  3.031134 -0.132642  0.158530
40  control_21 -1.197505   74.020866  0.440664  1.093153  2.041894 -0.774692
41  control_22 -0.075306   -3.122585  1.550893  0.676102  0.040827 -2.029262
42  control_23  0.076952   -6.802446 -1.232931 -0.018436 -2.433636  1.050102
43  control_24  0.614004  -20.279352 -0.516627 -2.199014 -3.231893 -0.931285
44  control_25  0.344809   23.927681  0.624941 -0.795765 -1.189744  1.352253
45  control_26  0.083489   83.022682  1.550425  1.399600  1.354261  3.687218
46  control_27  0.123651   67.629562 -0.086028  0.524495  0.971809  0.548718
47  control_28 -0.058165   25.141212  1.209469  1.359949  0.507893  0.976440
48  control_29 -0.236342   71.019707 -0.462640  3.749186 -1.277605  0.425622
49   control_3  0.167366  -34.797283  0.357245 -0.066158  1.168891  0.177882
50  control_30 -0.643911  -14.000072  1.835436  1.428437 -0.043603 -0.616939
51   control_4 -0.209942  122.459976 -0.920861  2.636191 -0.467058  1.950020
52   control_5  0.216940   70.276794 -0.047402 -0.267109  0.071883  0.603828
53   control_6 -0.079475   40.371067 -0.507461 -0.477919 -1.543805 -0.358059
54   control_7  0.432535   84.358543  0.949899  1.760492 -2.822368  1.914291
55   control_8 -0.063806   14.231878 -1.445765 -1.476629 -1.507881 -0.153884
56   control_9  0.369962  -66.017410 -1.536271 -1.589195 -1.286143 -2.146675

          7          8         9  …      2035      2037       2038  \
0  -1.131751 -14.571426 -1.861321  …  0.406672  0.184906  25.807522
1  -2.205090 -12.198885 -0.352454  … -1.560328  0.717220  -3.868195
2  -3.552037  -1.582109 -0.670278  … -2.544272 -0.683314  18.354477
3  -1.586803  -8.690717  0.951084  … -1.227247  0.302983 -12.904033
4  -4.464425 -19.646568  0.079340  … -4.855191 -1.832383  -6.097253
5  -2.822525 -11.954848  1.077436  … -0.002535  0.119448   0.092159
6   0.287346  -1.421384 -1.366856  …  3.441972  1.117718  15.763485
7   0.647401  -3.597660 -0.897405  …  2.745557 -2.432338   3.358941
8  -1.954207  -5.068198  1.349519  … -3.706210  0.801929  -6.272098
9  -1.094090   0.861966  0.347395  … -2.409920  0.340438   9.668518
10 -1.136406  -6.359642  0.955487  … -1.318173  0.332667 -14.074587
11 -2.096948  -8.931929  0.938213  … -1.724004  0.297168  -1.711162
12 -0.638754   0.266889  1.311957  … -0.475437  0.390245  -6.970936
13 -1.978605   0.298321 -0.181288  … -0.988314  0.346313 -13.934760
```

```
14  2.989360   5.116446  0.592651  …  1.114378  0.228223  10.715270
15  1.273390   4.367759  0.654822  … -0.701211  0.804361 -19.601755
16 -1.055286  -8.479153  0.164643  …  1.049131  0.323292  -9.187586
17 -1.349630   0.653174 -0.280863  … -1.653756 -0.849220   0.708098
18 -0.287695   7.340446 -1.620130  … -2.325716 -1.050491  14.312077
19 -4.603915  -3.069319  2.413633  … -3.113409  1.228905 -21.041458
20  2.987509   7.931948  0.137618  …  0.291863  0.438107  12.622253
21 -2.747367 -12.064553 -0.692529  …  1.308026  0.353445   3.849181
22  0.061767   2.212223 -0.322633  … -1.878157  1.129786  -9.621005
23  1.353153  -8.992580  0.619826  …  3.365197  0.286943   0.016523
24  3.915261   1.297010 -0.122544  … -0.060711  0.241144  -2.035032
25 -1.047273 -12.026852  2.107308  …  1.214462 -1.584782   0.239402
26  3.639369   2.065905 -1.196867  …  1.829335  0.319605  -5.941478
27  1.133942   4.150527 -2.790581  … -1.494848  0.102815   1.355160
28 -3.869670  -2.830388  0.565865  … -1.215553  0.594764  11.332007
29  4.664620   9.840240 -0.973511  …  1.617748 -0.410882  18.159657
30 -2.905175   5.176883 -0.289854  … -1.215137 -0.271957  -8.111947
31 -0.548172  11.652799 -1.127933  … -0.223871  0.259768  14.684847
32  1.793109   5.839436 -0.598026  …  2.084352  0.397427  -1.600280
33  2.242677   9.566770 -0.238075  …  2.165102 -1.733388  -3.674297
34  0.443357   8.664005  0.374799  …  1.679695 -1.052620  12.478277
35 -0.505418   9.848477 -0.622583  …  4.341630 -2.870811   9.249910
36 -1.239613   2.651411  0.088271  … -0.131776  0.288727  12.987974
37  0.076592   0.642919  0.329105  … -1.717374  1.328499 -18.217957
38  2.329894   6.439797  0.953147  …  1.983269  1.065727  -0.180908
39  2.009258   3.419312  1.274807  … -2.272105  0.287751 -11.898565
40 -0.541657   3.600390 -0.125512  … -1.326065  0.299974   9.117936
41 -3.878952  -2.323802  0.316967  … -1.823567  0.296469  -0.920544
42  3.309128  -7.326036  0.973565  … -0.538256  0.268268 -18.922113
43 -0.180823  -3.996730 -0.335466  … -0.466858  1.121918 -13.139399
44  0.900903  -0.020428  0.150159  …  0.494123 -0.002509  -0.907416
45  1.650264   4.742365 -0.448791  …  3.536846  0.861965   3.117191
46 -0.135439   2.961195 -0.316009  …  0.179191 -2.023907   9.100789
47  1.485703   0.743416 -0.080703  …  2.448230  0.497622 -10.821836
48  3.277027  12.077893 -1.521274  …  5.266572  1.193347  11.366472
49 -0.444897  -9.310042  1.136533  … -2.724898 -0.588951  -5.957805
50 -2.275975   3.176654  1.169152  …  1.275269  0.221227  -5.287311
51  5.771136   8.284544 -2.354664  …  2.100912  0.330918   9.033189
52 -0.313221   5.947603 -0.313333  …  0.414689 -1.278962   0.152494
53  2.865170   6.012676 -0.227228  …  2.048009  1.178996   8.447434
54  1.292099   3.270333  0.419287  …  0.756579  0.294694   9.825337
55  1.537149   4.266818  0.081452  … -0.996916  0.243656  -7.869002
56 -1.344763  -0.925296  0.394672  … -2.466995 -2.772863 -15.145872

        2039      2040      2041      2042      2044      2045  label
0  -1.525499  0.619345 -7.009365  0.480932 -1.841229 -0.770367      1
1  -2.123830  0.118527 -2.376938 -1.840528  1.806760 -0.984125      1
```

2	-1.892688	1.364370	-1.210879	-2.946210	-0.358407	-1.337365	1
3	-0.696032	1.509288	-0.605615	-3.349327	-1.091852	3.288409	1
4	-2.021634	0.362182	-2.012872	-2.203544	-0.237049	1.347519	1
5	-4.960733	-0.083534	-2.309518	-1.950238	2.953205	-0.614862	1
6	-1.848888	-0.602811	1.913821	2.543114	-0.462794	0.700600	1
7	0.669057	-0.387292	0.790129	0.113250	0.291465	-0.135850	1
8	-1.540754	0.429791	0.338263	0.375218	2.936503	0.850150	1
9	-0.080603	-0.711015	1.650100	0.728260	2.203407	-0.558213	1
10	-1.277359	1.277176	-1.235959	-2.874829	0.163668	-0.355272	1
11	-1.238571	-1.735411	-2.298757	-0.509185	0.803254	0.907511	1
12	-0.176494	0.430384	2.085759	0.469348	-1.290333	1.220300	1
13	1.554878	2.522246	1.685867	0.119645	-1.331253	0.316728	1
14	0.635167	0.232898	-0.065892	2.921533	-2.556237	-0.022152	1
15	-1.706216	-0.866527	-1.106549	1.781245	-1.129261	0.852984	1
16	1.954600	0.345905	4.136502	1.461302	-1.452394	0.439057	1
17	-1.406832	0.159083	0.840393	-1.944591	-1.460755	-0.643267	1
18	1.251680	-1.175103	0.569067	-1.122405	0.257440	0.719633	1
19	-7.307338	-2.194023	-3.254924	-1.988927	1.917969	-1.890589	1
20	-0.357630	1.779748	1.993150	3.449870	-0.382382	-0.440085	1
21	0.942220	0.888677	1.895064	-1.099630	-0.327086	-0.486317	1
22	-0.890594	0.457218	-0.183446	0.798839	-0.783525	0.212189	1
23	4.062076	0.410181	1.277490	0.719003	3.229269	-1.079710	1
24	0.204375	-1.489290	-2.365291	-3.476831	0.716498	-0.414338	1
25	-0.901431	-1.161160	-1.152193	-2.109130	1.458242	-2.580553	1
26	5.130542	-1.382279	2.854811	2.408052	-1.091843	1.390827	1
27	0.047281	0.134717	-0.790449	0.424735	-2.352363	-0.276436	0
28	0.621902	0.878529	-1.484326	0.168629	-0.103705	-0.003821	0
29	0.566339	-0.458097	-0.682318	1.207550	-0.998093	0.505997	0
30	0.117637	0.985556	0.750336	-0.013471	0.633004	1.060287	0
31	-0.229282	-0.478021	-2.262308	1.946468	-0.017038	1.137515	0
32	1.142318	0.103366	0.422751	0.951039	-2.661158	-0.100720	0
33	3.744782	-0.476126	2.381333	1.317338	0.248268	0.575689	0
34	1.997771	-0.275007	0.079734	0.163391	1.736207	-0.117855	0
35	0.897067	-0.797411	0.910178	3.183214	-1.173505	0.227411	0
36	-0.550152	0.161523	-2.619191	-0.353035	2.970695	0.485959	0
37	-0.392360	2.289821	-2.208766	-3.863082	-3.181592	0.165411	0
38	2.191755	-0.287665	0.660800	-0.481297	2.028596	-1.726705	0
39	-2.253137	-0.829837	0.200065	-2.363508	2.313675	-1.387579	0
40	-0.972172	-1.061687	-2.142606	0.858676	0.759539	-0.412720	0
41	-4.051919	-0.776241	-1.162009	0.594886	-0.142280	0.560845	0
42	-1.099188	-1.249040	-0.929002	0.148090	-0.487398	-2.852219	0
43	0.539770	1.070168	-0.348630	-1.707103	-1.946502	0.812911	0
44	0.755133	0.438451	-1.577958	-0.045895	-0.018299	-0.615549	0
45	1.734670	-0.018837	0.547540	2.848412	-0.739127	1.240845	0
46	-0.357706	0.084153	-1.441937	-0.167291	0.253577	0.254834	0
47	0.539878	-0.411115	6.645276	0.346189	0.784649	0.165318	0
48	3.746002	-2.211902	1.463393	2.688958	1.834320	0.107379	0

```
49 -1.440679  0.245530 -2.717933 -0.090637  0.667261  0.008718      0
50 -0.279829 -1.259002  3.082096  0.110309 -0.340987 -0.506620      0
51  2.811098  0.042105  4.512824  0.704178 -0.403841  0.372538      0
52  0.093633  0.520501 -0.827331  0.248041 -0.144364  0.272101      0
53  3.635352  0.871489  6.499184  0.907421 -1.836129  0.560169      0
54  2.353171  0.327373 -1.450833 -0.053466  0.773353  0.130662      0
55 -0.200318  1.126032 -1.710859  0.285345 -0.635436 -0.541088      0
56 -0.160282  0.192101  1.358729 -0.918320 -0.762609 -0.036119      0

[57 rows x 1618 columns]
```

[21]:

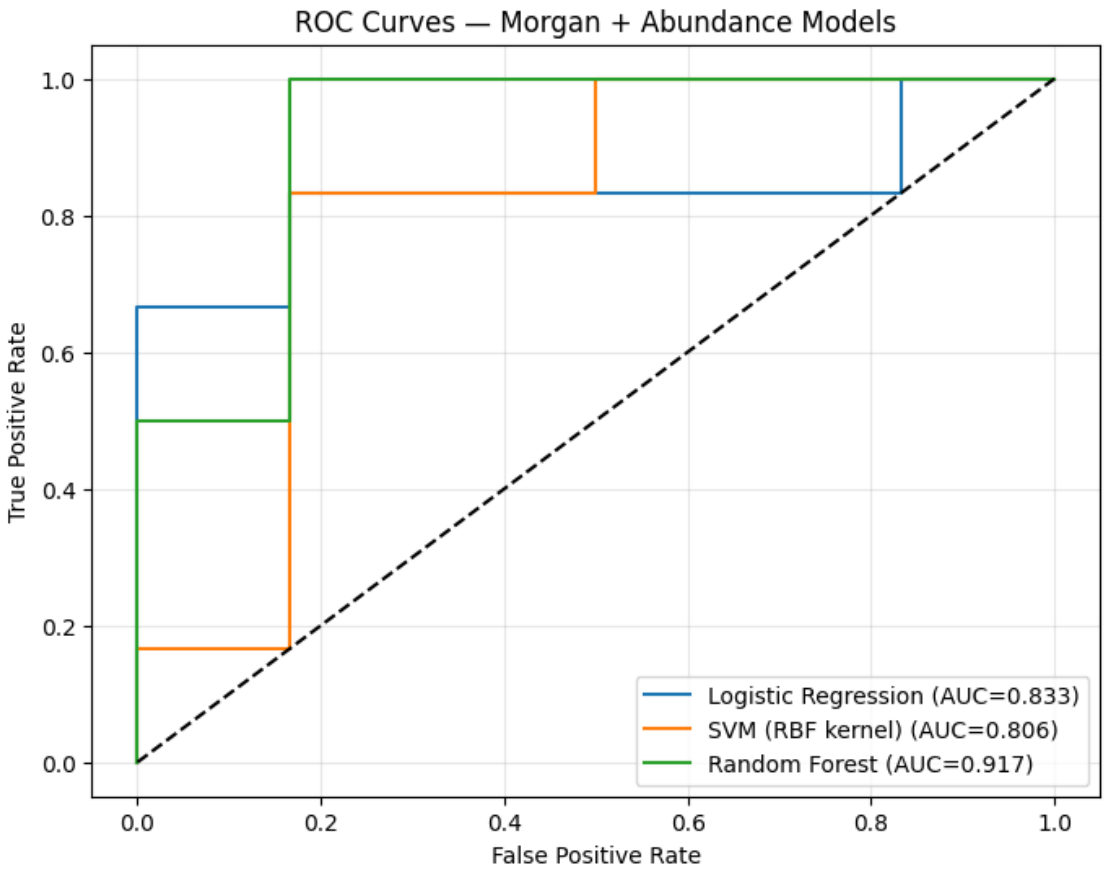
```python
# ============================
#  ML Baseline: Morgan + Abundance
# ============================

import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.metrics import roc_auc_score, roc_curve, confusion_matrix,␣
 ↪classification_report
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from sklearn.ensemble import RandomForestClassifier
import matplotlib.pyplot as plt
import seaborn as sns

# -------------------------------------
# 0. Define X, y (FROM morgan_abundance_ML)
# -------------------------------------
# morgan_abundance_ML columns:
# ['sample', feature_0, ..., feature_2045, 'label']

X = morgan_abundance_ML.drop(columns=["sample", "label"]).values
y = morgan_abundance_ML["label"].values

print("X shape:", X.shape)
print("y shape:", y.shape)

# -------------------------------------
# 1. Train-test split
# -------------------------------------
X_train, X_test, y_train, y_test = train_test_split(
    X,
    y,
    test_size=0.2,
    random_state=42,
    stratify=y
```

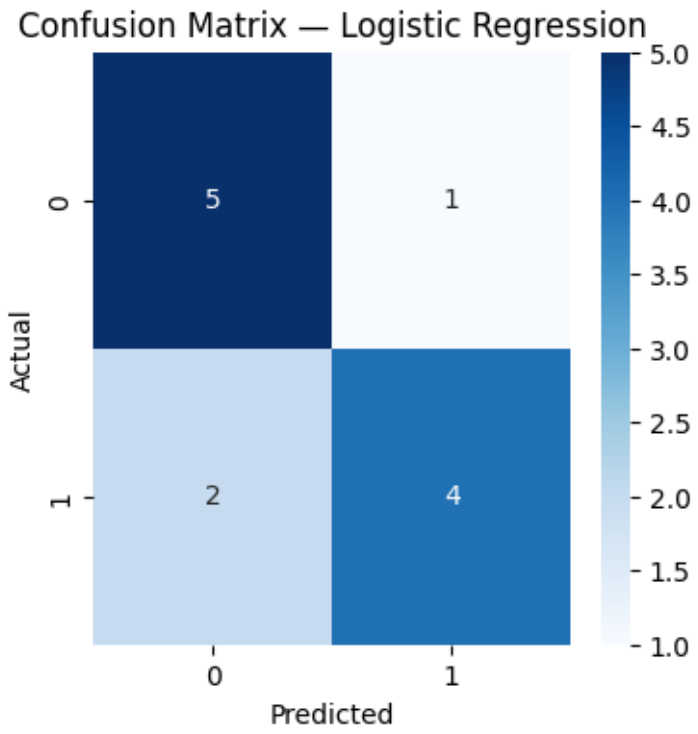
```
)

# ------------------------------------
# 2. Define models
# ------------------------------------
models = {
    "Logistic Regression": LogisticRegression(max_iter=500),
    "SVM (RBF kernel)": SVC(kernel="rbf", probability=True),
    "Random Forest": RandomForestClassifier(
        n_estimators=200,
        random_state=42
    )
}

results = {}

# ------------------------------------
# 3. Train & Evaluate models
# ------------------------------------
for name, model in models.items():
    print(f"\n==============================")
    print(f"Training: {name}")
    print("==============================")

    # Train
    model.fit(X_train, y_train)

    # Predict probability for ROC
    y_prob = model.predict_proba(X_test)[:, 1]

    # Metrics
    auc = roc_auc_score(y_test, y_prob)
    y_pred = model.predict(X_test)

    results[name] = {
        "model": model,
        "auc": auc,
        "confusion_matrix": confusion_matrix(y_test, y_pred),
        "classification_report": classification_report(
            y_test, y_pred, output_dict=False
        )
    }

    print(f"AUC: {auc:.4f}")
    print(results[name]["classification_report"])

# ------------------------------------
```

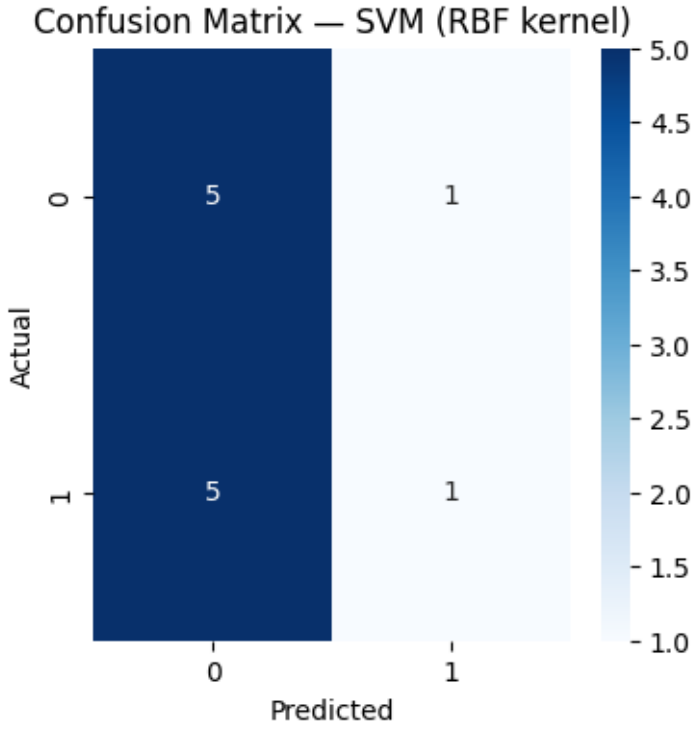
```python
# 4. Plot ROC curves
# -------------------------------------
plt.figure(figsize=(8, 6))

for name, result in results.items():
    model = result["model"]
    y_prob = model.predict_proba(X_test)[:, 1]
    fpr, tpr, _ = roc_curve(y_test, y_prob)

    plt.plot(
        fpr,
        tpr,
        label=f"{name} (AUC={result['auc']:.3f})"
    )

plt.plot([0, 1], [0, 1], "k--")
plt.xlabel("False Positive Rate")
plt.ylabel("True Positive Rate")
plt.title("ROC Curves - Morgan + Abundance Models")
plt.legend()
plt.grid(alpha=0.3)
plt.show()

# -------------------------------------
# 5. Plot confusion matrices
# -------------------------------------
for name, result in results.items():
    cm = result["confusion_matrix"]

    plt.figure(figsize=(4, 4))
    sns.heatmap(cm, annot=True, fmt="d", cmap="Blues")
    plt.title(f"Confusion Matrix - {name}")
    plt.xlabel("Predicted")
    plt.ylabel("Actual")
    plt.show()

# -------------------------------------
# 6. Cross-validation (5-fold)
# -------------------------------------
print("\n\n===== 5-fold Cross Validation (Morgan + Abundance) =====")

for name, model in models.items():
    cv_auc = cross_val_score(
        model,
        X,
        y,
        cv=5,
```

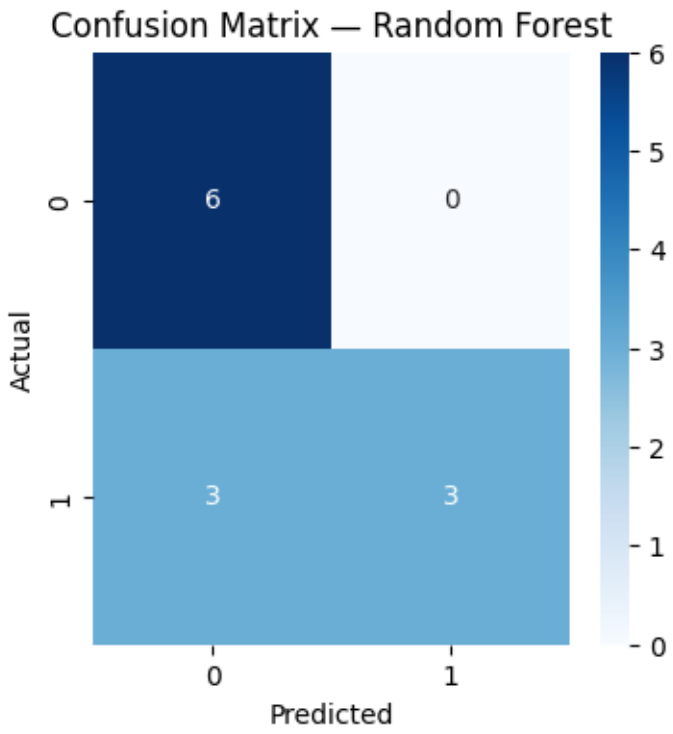
```
        scoring="roc_auc"
    )
    print(f"{name} CV AUC: mean={cv_auc.mean():.4f}, std={cv_auc.std():.4f}")
```

```
X shape: (57, 1616)
y shape: (57,)

==============================
Training: Logistic Regression
==============================
AUC: 0.8333
              precision    recall  f1-score   support

           0       0.71      0.83      0.77         6
           1       0.80      0.67      0.73         6

    accuracy                           0.75        12
   macro avg       0.76      0.75      0.75        12
weighted avg       0.76      0.75      0.75        12


==============================
Training: SVM (RBF kernel)
==============================
AUC: 0.8056
              precision    recall  f1-score   support

           0       0.50      0.83      0.62         6
           1       0.50      0.17      0.25         6

    accuracy                           0.50        12
   macro avg       0.50      0.50      0.44        12
weighted avg       0.50      0.50      0.44        12


==============================
Training: Random Forest
==============================
AUC: 0.9167
              precision    recall  f1-score   support

           0       0.67      1.00      0.80         6
           1       1.00      0.50      0.67         6

    accuracy                           0.75        12
   macro avg       0.83      0.75      0.73        12
weighted avg       0.83      0.75      0.73        12
```


ROC Curves — Morgan + Abundance Models
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Logistic Regression (AUC=0.833)
SVM (RBF kernel) (AUC=0.806)
Random Forest (AUC=0.917)


Confusion Matrix — Logistic Regression
Actual
Predicted
0
1
5
1
2
4
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0


Confusion Matrix — SVM (RBF kernel)
Actual
Predicted
0
1
5
1
5
1
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0




```
===== 5-fold Cross Validation (Morgan + Abundance) =====
Logistic Regression CV AUC: mean=0.8989, std=0.0731
SVM (RBF kernel) CV AUC: mean=0.8111, std=0.0916
Random Forest CV AUC: mean=0.9011, std=0.0861
```

[ ]: