SUPPLEMENTARY MATERIAL
Restricted Set Classification with prior probabilities: A case study on chessboard recognition

Ludmila I. Kuncheva and James H. V. Constance

(under revision for *Pattern Recognition Letters*)

Description of the data collection and organisation

The Week in Chess Magazine[1] publishes a weekly database of games from recent tournaments and matches by players of all skill levels, from amateur to the best in the world. Edition 1144, published on 10/10/2016 and including 3583 games, was downloaded in Portable Game Notation (PGN) format as the basis for a dataset of positions. The average FIDE (Fédération Internationale des Échecs) Elo rating of the players in the database was approximately 2300, which corresponds to the strength of a FIDE Master, indicating that the games are of a good quality. It was assumed that the games are representative of rated games played during any particular week during the year.

The PGN file was converted to EPD (Extended Position Description) format using pgn-extract[2]. EPD files are text files, in which each position is summarised on one line by reading the order of pieces and spaces from a board, row by row. Figure 1 shows an example of a position and how it is represented as an EPD string. The resulting EPD file contained 310666 positions from the games.

A MATLAB function was written to translate the EPD file into a dataset of positions (henceforth known as the 'EPD dataset'), with each row consisting of the occupancy of each of the 64 squares on the board. The occupancy could be one of 13 classes, these being the 6 types of piece (king, queen, rook, bishop, knight and pawn) for white and black, and empty squares as shown in Table 1.

Each row of 64 squares represents one board read from left to right and from top to bottom. Thus the order of the squares in the row is:

```
a8  b8  c8  d8  e8  f8  g8  h8  a7  b7  ...
```

[1] M. Crowther, "The week in chess", `http://theweekinchess.com/twic`

[2] D. J. Barnes,"pgn-extract: A portable game notation (pgn) manipulator for chess games," `https://www.cs.kent.ac.uk/people/staff/djb/pgn-extract/`

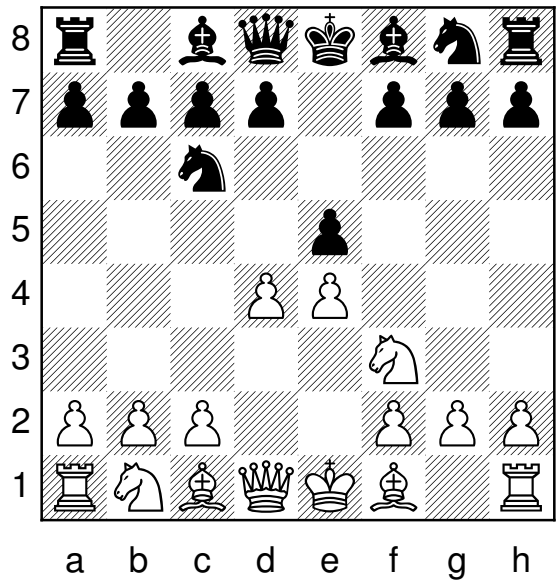

Figure 1: Position represented by the EPD string r1bqkbnr/pppp1ppp/2n5/4p3/3PP3/5N2/PPP2PPP/RNBQKB1R, where small letters represent black pieces, capital letters white pieces, and numbers the empty squares.

Table 1: Classes in the chess-pieces recognition problem, and the limit number for each class in a standard chess game.

						Class #						
1	2	3	4	5	6	7	8	9	10	11	12	13
												
king	queen	rook	bishop	knight	pawn	king	queen	rook	bishop	knight	pawn	empty

As an example, consider the board in Figure 1. The first 20 entries in the row corresponding to this board would be as follows:

```
9  13  10  8  7  10  11  9  12  12  12  12  13  12  12  12  13  13  11  13  ...
```

A MATLAB file `chess_data_twic1144_positions.mat` is available in GitHub at `https://github.com/LucyKuncheva/Chess-piece-recognition`.